




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C07K 13/00, 15/28, A61K 37/36, C12N 15/18, 15/10, 15/66</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 94/21681</b> <b>(43) International Publication Date:</b> 29 September 1994 (29.09.94)
<b>(21) International Application Number:</b> PCT/US94/03019 <b>(22) International Filing Date:</b> 18 March 1994 (18.03.94)  <b>(30) Priority Data:</b> 08/033,923      19 March 1993 (19.03.93)      US  <b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> US      08/033,923 (CIP) Filed on      19 March 1993 (19.03.93)  <b>(71) Applicant (for all designated States except US):</b> JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE [US/US]; 720 Rutland Avenue, Baltimore, MD 21205 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> LEE, Se-Jin [US/US]; 6711 Chokeberry Road, Baltimore, MD 21209 (US). McPHERRON, Alexandra, C. [US/US]; 3905 Keswick Road, Baltimore, MD 21211 (US).  <b>(74) Agents:</b> WETHERELL, John, R. et al.; Spensley Horn Jubas & Lubitz, 1880 Century Park East, 5th Floor, Los Angeles, CA 90067 (US).		<b>(81) Designated States:</b> CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> GROWTH DIFFERENTIATION FACTOR-8  <b>(57) Abstract</b> <p>Growth differentiation factor-8 (GDF-8) is disclosed along with its polynucleotide sequence and amino acid sequence. Also disclosed are diagnostic and therapeutic methods of using the GDF-8 polypeptide and polynucleotide sequences.</p> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="flex: 1;">  </div> <div style="flex: 1; text-align: right;">           HEART            LUNG            THYMUS            BRAIN            KIDNEY            SEMINAL VESICLE            PANCREAS            INTESTINE            SPLEEN            TESTIS            MUSCLE            LIVER            OVARY            FAT            UTERUS         </div> </div> <div style="text-align: center; margin-top: 20px;">       - 2.9 kb     </div>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## **GROWTH DIFFERENTIATION FACTOR-8**

This application is a continuation-in-part application of the U.S. Application Serial No. 08/033,923 filed on 3/19/93.

### **BACKGROUND OF THE INVENTION**

#### **5        1.        *Field of the Invention***

The invention relates generally to growth factors and specifically to a new member of the transforming growth factor beta (TGF- $\beta$ ) superfamily, which is denoted, growth differentiation factor-8 (GDF-8).

#### **2.        *Description of Related Art***

10        The transforming growth factor  $\beta$  (TGF- $\beta$ ) superfamily encompasses a group of structurally-related proteins which affect a wide range of differentiation processes during embryonic development. The family includes, Mullerian inhibiting substance (MIS), which is required for normal male sex development (Behringer, et al., Nature, 345:167, 1990), Drosophila decapentaplegic (DPP)  
15        gene product, which is required for dorsal-ventral axis formation and morphogenesis of the imaginal disks (Padgett, et al., Nature, 325:81-84, 1987), the Xenopus Vg-1 gene product, which localizes to the vegetal pole of eggs ((Weeks, et al., Cell, 51:861-867, 1987), the activins (Mason, et al., Biochem. Biophys. Res. Commun., 135:957-964, 1986), which can induce the formation  
20        of mesoderm and anterior structures in Xenopus embryos (Thomsen, et al., Cell, 63:485, 1990), and the bone morphogenetic proteins (BMPs, osteogenin, OP-1) which can induce de novo cartilage and bone formation (Sampath, et al., J. Biol. Chem., 265:13198, 1990). The TGF- $\beta$ s can influence a variety of differentiation processes, including adipogenesis, myogenesis, chondrogenesis,

hematopoiesis, and epithelial cell differentiation (for review, see Massague, Cell 49:437, 1987).

The proteins of the TGF- $\beta$  family are initially synthesized as a large precursor protein which subsequently undergoes proteolytic cleavage at a cluster of basic residues approximately 110-140 amino acids from the C-terminus. The C-terminal regions, or mature regions, of the proteins are all structurally related and the different family members can be classified into distinct subgroups based on the extent of their homology. Although the homologies within particular subgroups range from 70% to 90% amino acid sequence identity, the homologies between subgroups are significantly lower, generally ranging from only 20% to 50%. In each case, the active species appears to be a disulfide-linked dimer of C-terminal fragments. Studies have shown that when the pro-region of a member of the TGF- $\beta$  family is coexpressed with a mature region of another member of the TGF- $\beta$  family, intracellular dimerization and secretion of biologically active homodimers occur (Gray, A., and Maston, A., Science, 247:1328, 1990). Additional studies by Hammonds, et al., (Molec. Endocrin. 5:149, 1991) showed that the use of the BMP-2 pro-region combined with the BMP-4 mature region led to dramatically improved expression of mature BMP-4. For most of the family members that have been studied, the homodimeric species has been found to be biologically active, but for other family members, like the inhibins (Ling, et al., Nature, 321:779, 1986) and the TGF- $\beta$ s (Cheifetz, et al., Cell, 48:409, 1987), heterodimers have also been detected, and these appear to have different biological properties than the respective homodimers.

Identification of new factors that are tissue-specific in their expression pattern will provide a greater understanding of that tissue's development and function.

### SUMMARY OF THE INVENTION

The present invention provides a cell growth and differentiation factor, GDF-8, a polynucleotide sequence which encodes the factor, and antibodies which are immunoreactive with the factor. This factor appears to relate to various cell proliferative disorders, especially those involving those involving muscle, nerve, and adipose tissue.

Thus, in one embodiment, the invention provides a method for detecting a cell proliferative disorder of muscle, nerve, or fat origin and which is associated with GDF-8. In another embodiment, the invention provides a method for treating a cell proliferative disorder by suppressing or enhancing GDF-8 activity.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a Northern blot showing expression of GDF-8 mRNA in adult tissues. The probe was a partial murine GDF-8 clone.

5      FIGURE 2 shows nucleotide and predicted amino acid sequences of murine GDF-8 (FIGURE 2a) and human GDF-8 (FIGURE 2b). The putative dibasic processing sites in the murine sequence are boxed.

FIGURE 3 shows the alignment of the C-terminal sequences of GDF-8 with other members of the TGF- $\beta$  superfamily. The conserved cysteine residues are boxed. Dashes denote gaps introduced in order to maximize alignment.

10      FIGURE 4 shows amino acid homologies among different members of the TGF- $\beta$  superfamily. Numbers represent percent amino acid identities between each pair calculated from the first conserved cysteine to the C-terminus. Boxes represent homologies among highly-related members within particular subgroups.

15      FIGURE 5 shows the sequence of GDF-8. Nucleotide and amino acid sequences of murine (FIGURE 5a) and human (FIGURE 5b) GDF-8 cDNA clones are shown. Numbers indicate nucleotide position relative to the 5' end. Consensus N-linked glycosylation signals are shaded. The putative RXXR proteolytic cleavage sites are boxed.

20      FIGURE 6 shows a hydropathicity profile of GDF-8. Average hydrophobicity values for murine (FIGURE 6a) and human (FIGURE 6b) GDF-8 were calculated using the method of Kyte and Doolittle (J. Mol. Biol., 157:105-132, 1982). Positive numbers indicate increasing hydrophobicity.

FIGURE 7 shows a comparison of murine and human GDF-8 amino acid sequences. The predicted murine sequence is shown in the top lines and the predicted human sequence is shown in the bottom lines. Numbers indicate amino acid position relative to the N-terminus. Identities between the two sequences are denoted by a vertical line.

FIGURE 8 shows the expression of GDF-8 in bacteria. BL21 (DE3) (pLysS) cells carrying a pRSET/GDF-8 expression plasmid were induced with isopropylthio- $\beta$ -galactoside, and the GDF-8 fusion protein was purified by metal chelate chromatography. Lanes: total=total cell lysate; soluble=soluble protein fraction; insoluble=insoluble protein fraction (resuspended in 10 mM Tris pH 8.0, 50 mM sodium phosphate, 8 M urea, and 10 mM  $\beta$ -mercaptoethanol [buffer B]) loaded onto the column; pellet=insoluble protein fraction discarded before loading the column; flowthrough=proteins not bound by the column; washes=washes carried out in buffer B at the indicated pH's. Positions of molecular weight standards are shown at the right. Arrow indicates the position of the GDF-8 fusion protein.

FIGURE 9 shows the expression of GDF-8 in mammalian cells. Chinese hamster ovary cells were transfected with pMSXND/GDF-8 expression plasmids and selected in G418. Conditioned media from G418-resistant cells (prepared from cells transfected with constructs in which GDF-8 was cloned in either the antisense or sense orientation) were concentrated, electrophoresed under reducing conditions, blotted, and probed with anti-GDF-8 antibodies and [ $^{125}$ I]iodoproteinA. Arrow indicates the position of the processed GDF-8 protein.

FIGURE 10 shows the expression of GDF-8 mRNA. Poly A-selected RNA (5  $\mu$ g each) prepared from adult tissues (FIGURE 10a) or placentas and embryos (FIGURE 10b) at the indicated days of gestation was electrophoresed on formaldehyde gels, blotted, and probed with full length murine GDF-8.

- 5      FIGURE 11 shows chromosomal mapping of human GDF-8. DNA samples prepared from human/rodent somatic cell hybrid lines were subjected to PCR, electrophoresed on agarose gels, blotted, and probed. The human chromosome contained in each of the hybrid cell lines is identified at the top of each of the first 24 lanes (1-22, X, and Y). In the lanes designated M, CHO, and H, the starting DNA template was total genomic DNA from mouse, hamster, and human sources, respectively. In the lane marked B1, no template DNA was used. Numbers at left indicate the mobilities of DNA standards.
- 10



## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a growth and differentiation factor, GDF-8 and a polynucleotide sequence encoding GDF-8. GDF-8 is expressed at highest levels in muscle and at lower levels in adipose tissue. In one embodiment, the invention provides a method for detection of a cell proliferative disorder of muscle, nerve, or fat origin which is associated with GDF-8 expression. In another embodiment, the invention provides a method for treating a cell proliferative disorder by using an agent which suppresses or enhances GDF-8 activity.

10 The TGF- $\beta$  superfamily consists of multifunctional polypeptides that control proliferation, differentiation, and other functions in many cell types. Many of the peptides have regulatory, both positive and negative, effects on other peptide growth factors. The structural homology between the GDF-8 protein of this invention and the members of the TGF- $\beta$  family, indicates that GDF-8 is a new member of the family of growth and differentiation factors. Based on the known activities of many of the other members, it can be expected that GDF-8 will also possess biological activities that will make it useful as a diagnostic and therapeutic reagent.

20 In particular, certain members of this superfamily have expression patterns or possess activities that relate to the function of the nervous system. For example, the inhibins and activins have been shown to be expressed in the brain (Meunier, et al., Proc. Natl. Acad. Sci., USA, 85:247, 1988; Sawchenko, et al., Nature, 334:615, 1988), and activin has been shown to be capable of functioning as a nerve cell survival molecule (Schubert, et al., Nature, 344:868, 25 1990). Another family member, namely, GDF-1, is nervous system-specific in its expression pattern (Lee, S.J., Proc. Natl. Acad. Sci., USA, 88:4250, 1991), and certain other family members, such as Vgr-1 (Lyons, et al., Proc. Natl.

Acad. Sci., USA, 86:4554, 1989; Jones, et al., Development, 111:531, 1991), OP-1 (Ozkaynak, et al., J. Biol. Chem., 267:25220, 1992), and BMP-4 (Jones, et al., Development, 111:531, 1991), are also known to be expressed in the nervous system. Because it is known that skeletal muscle produces a factor or factors that promote the survival of motor neurons (Brown, Trends Neurosci., 7:10, 1984), the expression of GDF-8 in muscle suggests that one activity of GDF-8 may be as a trophic factor for neurons. In this regard, GDF-8 may have applications in the treatment of neurodegenerative diseases, such as amyotrophic lateral sclerosis, or in maintaining cells or tissues in culture prior to transplantation.

GDF-8 may also have applications in treating disease processes involving muscle, such as in musculodegenerative diseases or in tissue repair due to trauma. In this regard, many other members of the TGF- $\beta$  family are also important mediators of tissue repair. TGF- $\beta$  has been shown to have marked effects on the formation of collagen and to cause a striking angiogenic response in the newborn mouse (Roberts, et al., Proc. Natl. Acad. Sci., USA 83:4167, 1986). TGF- $\beta$  has also been shown to inhibit the differentiation of myoblasts in culture (Massague, et al., Proc. Natl. Acad. Sci., USA 83:8206, 1986). Moreover, because myoblast cells may be used as a vehicle for delivering genes to muscle for gene therapy, the properties of GDF-8 could be exploited for maintaining cells prior to transplantation or for enhancing the efficiency of the fusion process.

The expression of GDF-8 in adipose tissue also raises the possibility of applications for GDF-8 in the treatment of obesity or of disorders related to abnormal proliferation of adipocytes. In this regard, TGF- $\beta$  has been shown to be a potent inhibitor of adipocyte differentiation in vitro (Ignotz and Massague, Proc. Natl. Acad. Sci., USA 82:8530, 1985).

5 The term "substantially pure" as used herein refers to GDF-8 which is substantially free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. One skilled in the art can purify GDF-8 using standard techniques for protein purification. The substantially pure polypeptide will yield a single major band on a non-reducing polyacrylamide gel. The purity of the GDF-8 polypeptide can also be determined by amino-terminal amino acid sequence analysis. GDF-8 polypeptide includes functional fragments of the polypeptide, as long as the activity of GDF-8 remains. Smaller peptides containing the biological activity of GDF-8 are included in the invention.

10 The invention provides polynucleotides encoding the GDF-8 protein. These polynucleotides include DNA, cDNA and RNA sequences which encode GDF-8. It is understood that all polynucleotides encoding all or a portion of GDF-8 are also included herein, as long as they encode a polypeptide with GDF-8 activity. Such polynucleotides include naturally occurring, synthetic, and intentionally  
15 manipulated polynucleotides. For example, GDF-8 polynucleotide may be subjected to site-directed mutagenesis. The polynucleotide sequence for GDF-8 also includes antisense sequences. The polynucleotides of the invention include sequences that are degenerate as a result of the genetic code. There are 20 natural amino acids, most of which are specified by more than one  
20 codon. Therefore, all degenerate nucleotide sequences are included in the invention as long as the amino acid sequence of GDF-8 polypeptide encoded by the nucleotide sequence is functionally unchanged.

Specifically disclosed herein is a genomic DNA sequence containing a portion of the GDF-8 gene. The sequence contains an open reading frame  
25 corresponding to the predicted C-terminal region of the GDF-8 precursor protein. The encoded polypeptide is predicted to contain two potential proteolytic processing sites (KR and RR). Cleavage of the precursor at the downstream site would generate a mature biologically active C-terminal

fragment of 109 amino acids with a predicted molecular weight of approximately 12,400. Also, disclosed are full length murine and human GDF-8 cDNA sequences. The murine pre-pro-GDF-8 protein is 376 amino acids in length, which is encoded by a 2676 base pair nucleotide sequence, beginning  
5 at nucleotide 104 and extending to a TGA stop codon at nucleotide 1232. The human GDF-8 protein is 375 amino acids and is encoded by a 2743 base pair sequence, with the open reading frame beginning at nucleotide 59 and extending to nucleotide 1184.

The C-terminal region of GDF-8 following the putative proteolytic processing  
10 site shows significant homology to the known members of the TGF- $\beta$  superfamily. The GDF-8 sequence contains most of the residues that are highly conserved in other family members (see FIGURE 3). Like the TGF- $\beta$ s and inhibin  $\beta$ s, GDF-8 contains an extra pair of cysteine residues in addition to the 7 cysteines found in virtually all other family members. Among the known  
15 family members, GDF-8 is most homologous to Vgr-1 (45% sequence identity) (see FIGURE 4).

Minor modifications of the recombinant GDF-8 primary amino acid sequence may result in proteins which have substantially equivalent activity as compared to the GDF-8 polypeptide described herein. Such modifications may be  
20 deliberate, as by site-directed mutagenesis, or may be spontaneous. All of the polypeptides produced by these modifications are included herein as long as the biological activity of GDF-8 still exists. Further, deletion of one or more amino acids can also result in a modification of the structure of the resultant molecule without significantly altering its biological activity. This can lead to the  
25 development of a smaller active molecule which would have broader utility. For example, one can remove amino or carboxy terminal amino acids which are not required for GDF-8 biological activity.

The nucleotide sequence encoding the GDF-8 polypeptide of the invention includes the disclosed sequence and conservative variations thereof. The term "conservative variation" as used herein denotes the replacement of an amino acid residue by another, biologically similar residue. Examples of conservative variations include the substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as the substitution of arginine for lysine, glutamic for aspartic acid, or glutamine for asparagine, and the like. The term "conservative variation" also includes the use of a substituted amino acid in place of an unsubstituted parent amino acid provided that antibodies raised to the substituted polypeptide also immunoreact with the unsubstituted polypeptide.

DNA sequences of the invention can be obtained by several methods. For example, the DNA can be isolated using hybridization techniques which are well known in the art. These include, but are not limited to: 1) hybridization of genomic or cDNA libraries with probes to detect homologous nucleotide sequences, 2) polymerase chain reaction (PCR) on genomic DNA or cDNA using primers capable of annealing to the DNA sequence of interest, and 3) antibody screening of expression libraries to detect cloned DNA fragments with shared structural features.

Preferably the GDF-8 polynucleotide of the invention is derived from a mammalian organism, and most preferably from a mouse, rat, or human. Screening procedures which rely on nucleic acid hybridization make it possible to isolate any gene sequence from any organism, provided the appropriate probe is available. Oligonucleotide probes, which correspond to a part of the sequence encoding the protein in question, can be synthesized chemically. This requires that short, oligopeptide stretches of amino acid sequence must be known. The DNA sequence encoding the protein can be deduced from the

genetic code, however, the degeneracy of the code must be taken into account. It is possible to perform a mixed addition reaction when the sequence is degenerate. This includes a heterogeneous mixture of denatured double-stranded DNA. For such screening, hybridization is preferably performed on either single-stranded DNA or denatured double-stranded DNA. Hybridization is particularly useful in the detection of cDNA clones derived from sources where an extremely low amount of mRNA sequences relating to the polypeptide of interest are present. In other words, by using stringent hybridization conditions directed to avoid non-specific binding, it is possible, for example, to allow the autoradiographic visualization of a specific cDNA clone by the hybridization of the target DNA to that single probe in the mixture which is its complete complement (Wallace, et al., Nucl. Acid Res., 9:879, 1981).

The development of specific DNA sequences encoding GDF-8 can also be obtained by: 1) isolation of double-stranded DNA sequences from the genomic DNA; 2) chemical manufacture of a DNA sequence to provide the necessary codons for the polypeptide of interest; and 3) in vitro synthesis of a double-stranded DNA sequence by reverse transcription of mRNA isolated from a eukaryotic donor cell. In the latter case, a double-stranded DNA complement of mRNA is eventually formed which is generally referred to as cDNA.

Of the three above-noted methods for developing specific DNA sequences for use in recombinant procedures, the isolation of genomic DNA isolates is the least common. This is especially true when it is desirable to obtain the microbial expression of mammalian polypeptides due to the presence of introns.

The synthesis of DNA sequences is frequently the method of choice when the entire sequence of amino acid residues of the desired polypeptide product is known. When the entire sequence of amino acid residues of the desired polypeptide is not known, the direct synthesis of DNA sequences is not possible and the method of choice is the synthesis of cDNA sequences. Among the standard procedures for isolating cDNA sequences of interest is the formation of plasmid- or phage-carrying cDNA libraries which are derived from reverse transcription of mRNA which is abundant in donor cells that have a high level of genetic expression. When used in combination with polymerase chain reaction technology, even rare expression products can be cloned. In those cases where significant portions of the amino acid sequence of the polypeptide are known, the production of labeled single or double-stranded DNA or RNA probe sequences duplicating a sequence putatively present in the target cDNA may be employed in DNA/DNA hybridization procedures which are carried out on cloned copies of the cDNA which have been denatured into a single-stranded form (Jay, et al., Nucl. Acid Res., 11:2325, 1983).

A cDNA expression library, such as lambda gt11, can be screened indirectly for GDF-8 peptides having at least one epitope, using antibodies specific for GDF-8. Such antibodies can be either polyclonally or monoclonally derived and used to detect expression product indicative of the presence of GDF-8 cDNA.

DNA sequences encoding GDF-8 can be expressed in vitro by DNA transfer into a suitable host cell. "Host cells" are cells in which a vector can be propagated and its DNA expressed. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. However, such progeny are included when the term "host cell" is used.

Methods of stable transfer, meaning that the foreign DNA is continuously maintained in the host, are known in the art.

In the present invention, the GDF-8 polynucleotide sequences may be inserted into a recombinant expression vector. The term "recombinant expression  
5 vector" refers to a plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of the GDF-8 genetic sequences. Such expression vectors contain a promoter sequence which facilitates the efficient transcription of the inserted genetic sequence of the host. The expression vector typically contains an origin of replication, a promoter, as  
10 well as specific genes which allow phenotypic selection of the transformed cells. Vectors suitable for use in the present invention include, but are not limited to the T7-based expression vector for expression in bacteria (Rosenberg, et al., Gene, 56:125, 1987), the pMSXND expression vector for expression in mammalian cells (Lee and Nathans, J. Biol. Chem., 263:3521,  
15 1988) and baculovirus-derived vectors for expression in insect cells. The DNA segment can be present in the vector operably linked to regulatory elements, for example, a promoter (e.g., T7, metallothionein I, or polyhedrin promoters).

Polynucleotide sequences encoding GDF-8 can be expressed in either prokaryotes or eukaryotes. Hosts can include microbial, yeast, insect and  
20 mammalian organisms. Methods of expressing DNA sequences having eukaryotic or viral sequences in prokaryotes are well known in the art. Biologically functional viral and plasmid DNA vectors capable of expression and replication in a host are known in the art. Such vectors are used to incorporate DNA sequences of the invention. Preferably, the mature C-terminal  
25 region of GDF-8 is expressed from a cDNA clone containing the entire coding sequence of GDF-8. Alternatively, the C-terminal portion of GDF-8 can be expressed as a fusion protein with the pro- region of another member of the TGF- $\beta$  family or co-expressed with another pro- region (see for example,



Hammonds, et al., Molec. Endocrin. 5:149, 1991; Gray, A., and Mason, A., Science, 247:1328, 1990).

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where  
5 the host is prokaryotic, such as E. coli, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the  $\text{CaCl}_2$  method using procedures well known in the art. Alternatively,  $\text{MgCl}_2$  or  $\text{RbCl}$  can be used. Transformation can also be performed after forming a protoplast of the host cell if desired.

10 When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be cotransformed with DNA sequences encoding the GDF-8 of the invention, and a second foreign  
15 DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein. (see for example, Eukaryotic Viral Vectors, Cold Spring Harbor Laboratory, Gluzman, ed., 1982).

20 Isolation and purification of microbial expressed polypeptide, or fragments thereof, provided by the invention, may be carried out by conventional means including preparative chromatography and immunological separations involving monoclonal or polyclonal antibodies.

The invention includes antibodies immunoreactive with GDF-8 polypeptide or  
25 functional fragments thereof. Antibody which consists essentially of pooled monoclonal antibodies with different epitopic specificities, as well as distinct

monoclonal antibody preparations are provided. Monoclonal antibodies are made from antigen containing fragments of the protein by methods well known to those skilled in the art (Kohler, et al., Nature, 256:495, 1975). The term antibody as used in this invention is meant to include intact molecules as well as fragments thereof, such as Fab and F(ab')<sub>2</sub>, which are capable of binding an epitopic determinant on GDF-8.

The term "cell-proliferative disorder" denotes malignant as well as non-malignant cell populations which often appear to differ from the surrounding tissue both morphologically and genotypically. Malignant cells (i.e. cancer) develop as a result of a multistep process. The GDF-8 polynucleotide that is an antisense molecule is useful in treating malignancies of the various organ systems, particularly, for example, cells in muscle or adipose tissue. Essentially, any disorder which is etiologically linked to altered expression of GDF-8 could be considered susceptible to treatment with a GDF-8 suppressing reagent. One such disorder is a malignant cell proliferative disorder, for example.

The invention provides a method for detecting a cell proliferative disorder of muscle or adipose tissue which comprises contacting an anti-GDF-8 antibody with a cell suspected of having a GDF-8 associated disorder and detecting binding to the antibody. The antibody reactive with GDF-8 is labeled with a compound which allows detection of binding to GDF-8. For purposes of the invention, an antibody specific for GDF-8 polypeptide may be used to detect the level of GDF-8 in biological fluids and tissues. Any specimen containing a detectable amount of antigen can be used. A preferred sample of this invention is muscle tissue. The level of GDF-8 in the suspect cell can be compared with the level in a normal cell to determine whether the subject has a GDF-8-associated cell proliferative disorder. Preferably the subject is human.

The antibodies of the invention can be used in any subject in which it is desirable to administer in vitro or in vivo immunodiagnosis or immunotherapy. The antibodies of the invention are suited for use, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the antibodies in these immunoassays can be detectably labeled in various ways. Examples of types of immunoassays which can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays which are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

The antibodies of the invention can be bound to many different carriers and used to detect the presence of an antigen comprising the polypeptide of the invention. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding antibodies, or will be able to ascertain such, using routine experimentation.

There are many different labels and methods of labeling known to those of ordinary skill in the art. Examples of the types of labels which can be used in the present invention include enzymes, radioisotopes, fluorescent compounds, colloidal metals, chemiluminescent compounds, phosphorescent compounds, and bioluminescent compounds. Those of ordinary skill in the art will know of

other suitable labels for binding to the antibody, or will be able to ascertain such, using routine experimentation.

Another technique which may also result in greater sensitivity consists of coupling the antibodies to low molecular weight haptens. These haptens can  
5 then be specifically detected by means of a second reaction. For example, it is common to use such haptens as biotin, which reacts with avidin, or dinitrophenyl, puridoxal, and fluorescein, which can react with specific anti-hapten antibodies.

10 In using the monoclonal antibodies of the invention for the in vivo detection of antigen, the detectably labeled antibody is given a dose which is diagnostically effective. The term "diagnostically effective" means that the amount of detectably labeled monoclonal antibody is administered in sufficient quantity to enable detection of the site having the antigen comprising a polypeptide of the invention for which the monoclonal antibodies are specific.

15 The concentration of detectably labeled monoclonal antibody which is administered should be sufficient such that the binding to those cells having the polypeptide is detectable compared to the background. Further, it is desirable that the detectably labeled monoclonal antibody be rapidly cleared from the circulatory system in order to give the best target-to-background  
20 signal ratio.

As a rule, the dosage of detectably labeled monoclonal antibody for in vivo diagnosis will vary depending on such factors as age, sex, and extent of disease of the individual. Such dosages may vary, for example, depending on whether multiple injections are given, antigenic burden, and other factors  
25 known to those of skill in the art.

For in vivo diagnostic imaging, the type of detection instrument available is a major factor in selecting a given radioisotope. The radioisotope chosen must have a type of decay which is detectable for a given type of instrument. Still another important factor in selecting a radioisotope for in vivo diagnosis is that deleterious radiation with respect to the host is minimized. Ideally, a radioisotope used for in vivo imaging will lack a particle emission, but produce a large number of photons in the 140-250 keV range, which may readily be detected by conventional gamma cameras.

For in vivo diagnosis radioisotopes may be bound to immunoglobulin either directly or indirectly by using an intermediate functional group. Intermediate functional groups which often are used to bind radioisotopes which exist as metallic ions to immunoglobulins are the bifunctional chelating agents such as diethylenetriaminepentaacetic acid (DTPA) and ethylenediaminetetraacetic acid (EDTA) and similar molecules. Typical examples of metallic ions which can be bound to the monoclonal antibodies of the invention are  $^{111}\text{In}$ ,  $^{97}\text{Ru}$ ,  $^{67}\text{Ga}$ ,  $^{68}\text{Ga}$ ,  $^{72}\text{As}$ ,  $^{89}\text{Zr}$ , and  $^{201}\text{Tl}$ .

The monoclonal antibodies of the invention can also be labeled with a paramagnetic isotope for purposes of in vivo diagnosis, as in magnetic resonance imaging (MRI) or electron spin resonance (ESR). In general, any conventional method for visualizing diagnostic imaging can be utilized. Usually gamma and positron emitting radioisotopes are used for camera imaging and paramagnetic isotopes for MRI. Elements which are particularly useful in such techniques include  $^{157}\text{Gd}$ ,  $^{55}\text{Mn}$ ,  $^{162}\text{Dy}$ ,  $^{52}\text{Cr}$ , and  $^{56}\text{Fe}$ .

The monoclonal antibodies of the invention can be used in vitro and in vivo to monitor the course of amelioration of a GDF-8-associated disease in a subject. Thus, for example, by measuring the increase or decrease in the number of cells expressing antigen comprising a polypeptide of the invention or changes

in the concentration of such antigen present in various body fluids, it would be possible to determine whether a particular therapeutic regimen aimed at ameliorating the GDF-8-associated disease is effective. The term "ameliorate" denotes a lessening of the detrimental effect of the GDF-8-associated disease  
5 in the subject receiving therapy.

The present invention identifies a nucleotide sequence that can be expressed in an altered manner as compared to expression in a normal cell, therefore it is possible to design appropriate therapeutic or diagnostic techniques directed to this sequence. Thus, where a cell-proliferative disorder is associated with  
10 the expression of GDF-8, nucleic acid sequences that interfere with GDF-8 expression at the translational level can be used. This approach utilizes, for example, antisense nucleic acid and ribozymes to block translation of a specific GDF-8 mRNA, either by masking that mRNA with an antisense nucleic acid or by cleaving it with a ribozyme. Such disorders include neurodegenerative  
15 diseases, for example.

Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule (Weintraub, Scientific American, 262:40, 1990). In the cell, the antisense nucleic acids hybridize to the corresponding mRNA, forming a double-stranded molecule. The antisense  
20 nucleic acids interfere with the translation of the mRNA, since the cell will not translate a mRNA that is double-stranded. Antisense oligomers of about 15 nucleotides are preferred, since they are easily synthesized and are less likely to cause problems than larger molecules when introduced into the target GDF-8-producing cell. The use of antisense methods to inhibit the in vitro  
25 translation of genes is well known in the art (Marcus-Sakura, Anal.Biochem., 172:289, 1988).

Ribozymes are RNA molecules possessing the ability to specifically cleave other single-stranded RNA in a manner analogous to DNA restriction endonucleases. Through the modification of nucleotide sequences which encode these RNAs, it is possible to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, J.Amer.Med. Assn., 260:3030, 1988). A major advantage of this approach is that, because they are sequence-specific, only mRNAs with particular sequences are inactivated.

There are two basic types of ribozymes namely, tetrahymena-type (Hasselhoff, Nature, 334:585, 1988) and "hammerhead"-type. Tetrahymena-type ribozymes recognize sequences which are four bases in length, while "hammerhead"-type ribozymes recognize base sequences 11-18 bases in length. The longer the recognition sequence, the greater the likelihood that the sequence will occur exclusively in the target mRNA species. Consequently, hammerhead-type ribozymes are preferable to tetrahymena-type ribozymes for inactivating a specific mRNA species and 18-based recognition sequences are preferable to shorter recognition sequences.

The present invention also provides gene therapy for the treatment of cell proliferative or immunologic disorders which are mediated by GDF-8 protein. Such therapy would achieve its therapeutic effect by introduction of the GDF-8 antisense polynucleotide into cells having the proliferative disorder. Delivery of antisense GDF-8 polynucleotide can be achieved using a recombinant expression vector such as a chimeric virus or a colloidal dispersion system. Especially preferred for therapeutic delivery of antisense sequences is the use of targeted liposomes.

Various viral vectors which can be utilized for gene therapy as taught herein include adenovirus, herpes virus, vaccinia, or, preferably, an RNA virus such as a retrovirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of retroviral vectors in which a single foreign gene can be inserted include, but are not limited to: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), and Rous Sarcoma Virus (RSV). A number of additional retroviral vectors can incorporate multiple genes. All of these vectors can transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated. By inserting a GDF-8 sequence of interest into the viral vector, along with another gene which encodes the ligand for a receptor on a specific target cell, for example, the vector is now target specific. Retroviral vectors can be made target specific by attaching, for example, a sugar, a glycolipid, or a protein. Preferred targeting is accomplished by using an antibody to target the retroviral vector. Those of skill in the art will know of, or can readily ascertain without undue experimentation, specific polynucleotide sequences which can be inserted into the retroviral genome or attached to a viral envelope to allow target specific delivery of the retroviral vector containing the GDF-8 antisense polynucleotide.

Since recombinant retroviruses are defective, they require assistance in order to produce infectious vector particles. This assistance can be provided, for example, by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. These plasmids are missing a nucleotide sequence which enables the packaging mechanism to recognize an RNA transcript for encapsidation. Helper cell lines which have deletions of the packaging signal include, but are not limited to  $\psi$ 2, PA317 and PA12, for example. These cell lines produce empty virions, since no genome is packaged. If a retroviral vector is introduced into such cells in which the packaging signal is intact, but



the structural genes are replaced by other genes of interest, the vector can be packaged and vector virion produced.

Alternatively, NIH 3T3 or other tissue culture cells can be directly transfected with plasmids encoding the retroviral structural genes gag, pol and env, by  
5 conventional calcium phosphate transfection. These cells are then transfected with the vector plasmid containing the genes of interest. The resulting cells release the retroviral vector into the culture medium.

Another targeted delivery system for GDF-8 antisense polynucleotides is a colloidal dispersion system. Colloidal dispersion systems include macromole-  
10 cule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. The preferred colloidal system of this invention is a liposome. Liposomes are artificial membrane vesicles which are useful as delivery vehicles in vitro and in vivo. It has been shown that large unilamellar vesicles (LUV), which range  
15 in size from 0.2-4.0  $\mu\text{m}$  can encapsulate a substantial percentage of an aqueous buffer containing large macromolecules. RNA, DNA and intact virions can be encapsulated within the aqueous interior and be delivered to cells in a biologically active form (Fraley, et al., Trends Biochem. Sci., 6:77, 1981). In addition to mammalian cells, liposomes have been used for delivery of  
20 polynucleotides in plant, yeast and bacterial cells. In order for a liposome to be an efficient gene transfer vehicle, the following characteristics should be present: (1) encapsulation of the genes of interest at high efficiency while not compromising their biological activity; (2) preferential and substantial binding to a target cell in comparison to non-target cells; (3) delivery of the aqueous  
25 contents of the vesicle to the target cell cytoplasm at high efficiency; and (4) accurate and effective expression of genetic information (Mannino, et al., Biotechniques, 6:682, 1988).

The composition of the liposome is usually a combination of phospholipids, particularly high-phase-transition-temperature phospholipids, usually in combination with steroids, especially cholesterol. Other phospholipids or other lipids may also be used. The physical characteristics of liposomes depend on pH, ionic strength, and the presence of divalent cations.

Examples of lipids useful in liposome production include phosphatidyl compounds, such as phosphatidylglycerol, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, sphingolipids, cerebrosides, and gangliosides. Particularly useful are diacylphosphatidylglycerols, where the lipid moiety contains from 14-18 carbon atoms, particularly from 16-18 carbon atoms, and is saturated. Illustrative phospholipids include egg phosphatidylcholine, dipalmitoylphosphatidylcholine and distearoylphosphatidylcholine.

The targeting of liposomes can be classified based on anatomical and mechanistic factors. Anatomical classification is based on the level of selectivity, for example, organ-specific, cell-specific, and organelle-specific. Mechanistic targeting can be distinguished based upon whether it is passive or active. Passive targeting utilizes the natural tendency of liposomes to distribute to cells of the reticulo-endothelial system (RES) in organs which contain sinusoidal capillaries. Active targeting, on the other hand, involves alteration of the liposome by coupling the liposome to a specific ligand such as a monoclonal antibody, sugar, glycolipid, or protein, or by changing the composition or size of the liposome in order to achieve targeting to organs and cell types other than the naturally occurring sites of localization.

The surface of the targeted delivery system may be modified in a variety of ways. In the case of a liposomal targeted delivery system, lipid groups can be incorporated into the lipid bilayer of the liposome in order to maintain the

targeting ligand in stable association with the liposomal bilayer. Various linking groups can be used for joining the lipid chains to the targeting ligand.

Due to the expression of GDF-8 in muscle and adipose tissue, there are a variety of applications using the polypeptide, polynucleotide, and antibodies of the invention, related to these tissues. Such applications include treatment of cell proliferative disorders involving these and other tissues, such as neural tissue. In addition, GDF-8 may be useful in various gene therapy procedures.

The data in Example 6 shows that the human GDF-8 gene is located on chromosome 2. By comparing the chromosomal location of GDF-8 with the map positions of various human disorders, it should be possible to determine whether mutations in the GDF-8 gene are involved in the etiology of human diseases. For example, an autosomal recessive form of juvenile amyotrophic lateral sclerosis has been shown to map to chromosome 2 (Hentati, et al., Neurology, 42 [Suppl.3]:201, 1992). More precise mapping of GDF-8 and analysis of DNA from these patients may indicate that GDF-8 is, in fact, the gene affected in this disease. In addition, GDF-8 is useful for distinguishing chromosome 2 from other chromosomes.

The following examples are intended to illustrate but not limit the invention. While they are typical of those that might be used, other procedures known to those skilled in the art may alternatively be used.

**EXAMPLE 1**  
**IDENTIFICATION AND ISOLATION OF A NOVEL**  
**TGF- $\beta$  FAMILY MEMBER**

To identify a new member of the TGF- $\beta$  superfamily, degenerate  
5 oligonucleotides were designed which corresponded to two conserved regions  
among the known family members: one region spanning the two tryptophan  
residues conserved in all family members except MIS and the other region  
spanning the invariant cysteine residues near the C-terminus. These primers  
were used for polymerase chain reactions on mouse genomic DNA followed  
10 by subcloning the PCR products using restriction sites placed at the 5' ends  
of the primers, picking individual E. coli colonies carrying these subcloned  
inserts, and using a combination of random sequencing and hybridization  
analysis to eliminate known members of the superfamily.

GDF-8 was identified from a mixture of PCR products obtained with the primers  
15 SJL141: 5'-CCGGAATTCGGITGG(G/C/A)A(G/A/T/C)(A/G)A(T/C)TGG(A/G)TI  
(A/G)TI(T/G)CICC-3' (SEQ ID NO:1)  
SJL147: 5'-CCGGAATTC(G/A)CAI(G/C)C(G/A)CA(G/A)CT(G/A/T/C)  
TCIACI(G/A)(T/C)CAT-3' (SEQ ID NO:2)

PCR using these primers was carried out with 2  $\mu$ g mouse genomic DNA at  
20 94°C for 1 min, 50°C for 2 min, and 72°C for 2 min for 40 cycles.

PCR products of approximately 280 bp were gel-purified, digested with Eco RI,  
gel-purified again, and subcloned in the Bluescript vector (Stratagene, San  
Diego, CA). Bacterial colonies carrying individual subclones were picked into  
96 well microtiter plates, and multiple replicas were prepared by plating the  
25 cells onto nitrocellulose. The replicate filters were hybridized to probes

-27-

representing known members of the family, and DNA was prepared from non-hybridizing colonies for sequence analysis.

The primer combination of SJL141 and SJL147, encoding the amino acid sequences GW(H/Q/N/K/D/E)(D/N)W(V/I/M)(V/I/M)(A/S)P (SEQ ID NO:9) and  
5 M(V/I/M/T/A)V(D/E)SC(G/A)C (SEQ ID NO:10), respectively, yielded four previously identified sequences (BMP-4, inhibin  $\beta$ B, GDF-3 and GDF-5) and one novel sequence, which was designated GDF-8, among 110 subclones analyzed.

Human GDF-8 was isolated using the primers:

10 ACM13: 5'-CGCGGATCCAGAAGTCAAGGTGACAGACACAC-3' (SEQ ID NO:3);  
and  
ACM14: 5'-CGCGGATCCTCCTCATGAGCACCCACAGCGGTC-3' (SEQ ID NO:4)

PCR using these primers was carried out with one  $\mu$ g human genomic DNA at 94°C for 1 min, 58°C for 2 min, and 72°C for 2 min for 30 cycles. The PCR  
15 product was digested with Bam HI, gel-purified, and subcloned in the Bluescript vector (Stratagene, San Francisco, CA).

## **EXAMPLE 2**

### **EXPRESSION PATTERN AND SEQUENCE OF GDF-8**

To determine the expression pattern of GDF-8, RNA samples prepared from  
20 a variety of adult tissues were screened by Northern analysis. RNA isolation and Northern analysis were carried out as described previously (Lee, S.-J., Mol. Endocrinol., 4:1034, 1990) except that hybridization was carried out in 5X SSPE, 10% dextran sulfate, 50% formamide, 1% SDS, 200  $\mu$ g/ml salmon DNA, and 0.1% each of bovine serum albumin, ficoll, and polyvinylpyrrolidone. Five

micrograms of twice poly A-selected RNA prepared from each tissue (except for muscle, for which only 2  $\mu$ g RNA was used) were electrophoresed on formaldehyde gels, blotted, and probed with GDF-8. As shown in FIGURE 1, the GDF-8 probe detected a single mRNA species expressed at highest levels in muscle and at significantly lower levels in adipose tissue.

To obtain a larger segment of the GDF-8 gene, a mouse genomic library was screened with a probe derived from the GDF-8 PCR product. The partial sequence of a GDF-8 genomic clone is shown in FIGURE 2a. The sequence contains an open reading frame corresponding to the predicted C-terminal region of the GDF-8 precursor protein. The predicted GDF-8 sequence contains two potential proteolytic processing sites, which are boxed. Cleavage of the precursor at the second of these sites would generate a mature C-terminal fragment 109 amino acids in length with a predicted molecular weight of 12,400. The partial sequence of human GDF-8 is shown in FIGURE 2b. Assuming no PCR-induced errors during the isolation of the human clone, the human and mouse amino acid sequences in this region are 100% identical.

The C-terminal region of GDF-8 following the putative proteolytic processing site shows significant homology to the known members of the TGF- $\beta$  superfamily (FIGURE 3). FIGURE 3 shows the alignment of the C-terminal sequences of GDF-8 with the corresponding regions of human GDF-1 (Lee, Proc. Natl. Acad. Sci. USA, 88:4250-4254, 1991), human BMP-2 and 4 (Wozney, et al., Science, 242:1528-1534, 1988), human Vgr-1 (Celeste, et al., Proc. Natl. Acad. Sci. USA, 87:9843-9847, 1990), human OP-1 (Ozkaynak, et al., EMBO J., 9:2085-2093, 1990), human BMP-5 (Celeste, et al., Proc. Natl. Acad. Sci. USA, 87:9843-9847, 1990), human BMP-3 (Wozney, et al., Science, 242:1528-1534, 1988), human MIS (Cate, et al., Cell, 45:685-698, 1986), human inhibin alpha,  $\beta$ A, and  $\beta$ B (Mason, et al., Biochem. Biophys. Res. Commun., 135:957-964, 1986), human TGF- $\beta$ 1 (Derynck, et al., Nature, 316:701-705,

1985), humanTGF- $\beta$ 2 (deMartin, et al., EMBO J., 6:3673-3677, 1987), and human TGF- $\beta$ 3 (ten Dijke, et al., Proc. Natl. Acad. Sci. USA, 85:4715-4719, 1988). The conserved cysteine residues are boxed. Dashes denote gaps introduced in order to maximize the alignment.

- 5 GDF-8 contains most of the residues that are highly conserved in other family members, including the seven cysteine residues with their characteristic spacing. Like the TGF- $\beta$ s and inhibin  $\beta$ s, GDF-8 also contains two additional cysteine residues. In the case of TGF- $\beta$ 2, these two additional cysteine residues are known to form an intramolecular disulfide bond (Daopin, et al.,  
10 Science, 257:369, 1992; Schlunegger and Grutter, Nature, 358:430, 1992).

FIGURE 4 shows the amino acid homologies among the different members of the TGF- $\beta$  superfamily. Numbers represent percent amino acid identities between each pair calculated from the first conserved cysteine to the C-terminus. Boxes represent homologies among highly-related members within  
15 particular subgroups. In this region, GDF-8 is most homologous to Vgr-1 (45% sequence identity).

### **EXAMPLE 3**

#### **ISOLATION OF cDNA CLONES ENCODING MURINE AND HUMAN GDF-8**

In order to isolate full-length cDNA clones encoding murine and human GDF-8,  
20 cDNA libraries were prepared in the lambda ZAP II vector (Stratagene) using RNA prepared from skeletal muscle. From 5  $\mu$ g of twice poly A-selected RNA prepared from murine and human muscle, cDNA libraries consisting of 4.4 million and 1.9 million recombinant phage, respectively, were constructed according to the instructions provided by Stratagene. These libraries were  
25 screened without amplification. Library screening and characterization of cDNA

inserts were carried out as described previously (Lee, Mol. Endocrinol, 4:1034-1040).

From  $2.4 \times 10^6$  recombinant phage screened from the murine muscle cDNA library, greater than 280 positive phage were identified using a murine GDF-8 probe derived from a genomic clone, as described in Example 1. The entire nucleotide sequence of the longest cDNA insert analyzed is shown in FIGURE 5a and SEQ ID NO:11. The 2676 base pair sequence contains a single long open reading frame beginning with a methionine codon at nucleotide 104 and extending to a TGA stop codon at nucleotide 1232. Upstream of the putative initiating methionine codon is an in-frame stop codon at nucleotide 23. The predicted pre-pro-GDF-8 protein is 376 amino acids in length. The sequence contains a core of hydrophobic amino acids at the N-terminus suggestive of a signal peptide for secretion (FIGURE 6a), one potential N-glycosylation site at asparagine 72, a putative RXXR proteolytic cleavage site at amino acids 264-267, and a C-terminal region showing significant homology to the known members of the TGF- $\beta$  superfamily. Cleavage of the precursor protein at the putative RXXR site would generate a mature C-terminal GDF-8 fragment 109 amino acids in length with a predicted molecular weight of approximately 12,400.

From  $1.9 \times 10^6$  recombinant phage screened from the human muscle cDNA library, 4 positive phage were identified using a human GDF-8 probe derived by polymerase chain reaction on human genomic DNA. The entire nucleotide sequence of the longest cDNA insert is shown in FIGURE 5b and SEQ ID NO:13. The 2743 base pair sequence contains a single long open reading frame beginning with a methionine codon at nucleotide 59 and extending to a TGA stop codon at nucleotide 1184. The predicted pre-pro-GDF-8 protein is 375 amino acids in length. The sequence contains a core of hydrophobic amino acids at the N-terminus suggestive of a signal peptide for secretion



(FIGURE 6b), one potential N-glycosylation site at asparagine 71, and a putative RXXR proteolytic cleavage site at amino acids 263-266. FIGURE 7 shows a comparison of the predicted murine (top) and human (bottom) GDF-8 amino acid sequences. Numbers indicate amino acid position relative to the N-terminus. Identities between the two sequences are denoted by a vertical line. Murine and human GDF-8 are approximately 94% identical in the predicted pro-regions and 100% identical following the predicted RXXR cleavage sites.

**EXAMPLE 4**  
**PREPARATION OF ANTIBODIES AGAINST GDF-8 AND**  
**EXPRESSION OF GDF-8 IN MAMMALIAN CELLS**

In order to prepare antibodies against GDF-8, GDF-8 antigen was expressed  
5 as a fusion protein in bacteria. A portion of murine GDF-8 cDNA spanning  
amino acids 268-376 (mature region) was inserted into the pRSET vector  
(Invitrogen) such that the GDF-8 coding sequence was placed in frame with the  
initiating methionine codon present in the vector; the resulting construct  
created an open reading frame encoding a fusion protein with a molecular  
10 weight of approximately 16,600. The fusion construct was transformed into  
BL21 (DE3) (pLysS) cells, and expression of the fusion protein was induced by  
treatment with isopropylthio- $\beta$ -galactoside as described (Rosenberg, et al.,  
Gene, 56:125-135). The fusion protein was then purified by metal chelate  
chromatography according to the instructions provided by Invitrogen. A  
15 Coomassie blue-stained gel of unpurified and purified fusion proteins is shown  
in FIGURE 8.

The purified fusion protein was used to immunize both rabbits and chickens.  
Immunization of rabbits was carried out by Spring Valley Labs (Sykesville, MD),  
and immunization of chickens was carried out by HRP, Inc. (Denver, PA).  
20 Western analysis of sera both from immunized rabbits and from immunized  
chickens demonstrated the presence of antibodies directed against the fusion  
protein.

To express GDF-8 in mammalian cells, the murine GDF-8 cDNA sequence from  
nucleotides 48-1303 was cloned in both orientations downstream of the  
25 metallothionein I promoter in the pMSXND expression vector; this vector  
contains processing signals derived from SV40, a dihydrofolate reductase  
gene, and a gene conferring resistance to the antibiotic G418 (Lee and

Nathans, J. Biol. Chem., 263:3521-3527). The resulting constructs were transfected into Chinese hamster ovary cells, and stable tranfectants were selected in the presence of G418. Two milliliters of conditioned media prepared from the G418-resistant cells were dialyzed, lyophilized, electrophoresed under denaturing, reducing conditions, transferred to nitrocellulose, and incubated with anti-GDF-8 antibodies (described above) and [<sup>125</sup>I]iodoproteinA.

As shown in FIGURE 9, the rabbit GDF-8 antibodies (at a 1:500 dilution) detected a protein of approximately the predicted molecular weight for the mature C-terminal fragment of GDF-8 in the conditioned media of cells transfected with a construct in which GDF-8 had been cloned in the correct (sense) orientation with respect to the metallothionein promoter (lane 2); this band was not detected in a similar sample prepared from cells transfected with a control antisense construct (lane 1). Similar results were obtained using antibodies prepared in chickens. Hence, GDF-8 is secreted and proteolytically processed by these transfected mammalian cells.

### **EXAMPLE 5**

#### **EXPRESSION PATTERN OF GDF-8**

To determine the pattern of GDF-8, 5 µg of twice poly A-selected RNA prepared from a variety of murine tissue sources were subjected to Northern analysis. As shown in FIGURE 10a (and as shown previously in Example 2), the GDF-8 probe detected a single mRNA species present almost exclusively in skeletal muscle among a large number of adult tissues surveyed. On longer exposures of the same blot, significantly lower but detectable levels of GDF-8 mRNA were seen in fat, brain, thymus, heart, and lung. Hence, these results confirm the high degree of specificity of GDF-8 expression in skeletal muscle. GDF-8 mRNA was also detected in mouse embryos at both gestational ages

(day 12.5 and day 18.5 post-coital) examined but not in placentas at various stages of development (FIGURE 10b).

### **EXAMPLE 6**

#### **CHROMOSOMAL LOCALIZATION OF GDF-8**

5 In order to map the chromosomal location of GDF-8, DNA samples from human/rodent somatic cell hybrids (Drwina, et al., Genomics, 16:311-413, 1993; Dubois and Naylor, Genomics, 16:315-319, 1993) were analyzed by polymerase chain reaction followed by Southern blotting. Polymerase chain reaction was carried out using primer #83, 5'-  
10 CGCGGATCCGTGGATCTAAATGAGAACAGTGAGC-3' (SEQ ID NO:15) and primer #84, 5'-CGCGAATTCTCAGGTAATGATTGTTCCGTTGTAGCG-3'(SEQ ID NO:16) for 40 cycles at 94°C for 2 minutes, 60°C for 1 minute, and 72°C for 2 minutes. These primers correspond to nucleotides 119 to 143 (flanked by a Bam H1 recognition sequence), and nucleotides 394 to 418 (flanked by  
15 an Eco R1 recognition sequence), respectively, in the human GDF-8 cDNA sequence. PCR products were electrophoresed on agarose gels, blotted, and probed with oligonucleotide #100, 5'-ACACTAAATCTTCAAGAATA-3' (SEQ ID NO:17), which corresponds to a sequence internal to the region flanked by primer #83 and #84. Filters were hybridized in 6 X SSC, 1 X Denhardt's  
20 solution, 100µg/ml yeast transfer RNA, and 0.05% sodium pyrophosphate at 50°C.

As shown in FIGURE 11, the human-specific probe detected a band of the predicted size (approximately 320 base pairs) in the positive control sample (total human genomic DNA) and in a single DNA sample from the  
25 human/rodent hybrid panel. This positive signal corresponds to human chromosome 2. The human chromosome contained in each of the hybrid cell lines is identified at the top of each of the first 24 lanes (1-22, X, and Y). In the

lanes designated M, CHO, and H, the starting DNA template was total genomic DNA from mouse, hamster, and human sources, respectively. In the lane marked B1, no template DNA was used. Numbers at left indicate the mobilities of DNA standards. These data show that the human GDF-8 gene is located on chromosome 2.

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

**SUMMARY OF SEQUENCES**

SEQ ID NO: 1 is the nucleic acid sequence for clone SJL141.

SEQ ID NO: 2 is the nucleic acid sequence for clone SJL147.

SEQ ID NO: 3 is the nucleic acid sequence for clone ACM13.

5 SEQ ID NO: 4 is the nucleic acid sequence for clone ACM14.

SEQ ID NO: 5 is the partial nucleotide sequence and deduced amino acid sequence for murine GDF-8.

SEQ ID NO: 6 is the deduced partial amino acid sequence for murine GDF-8.

10 SEQ ID NO: 7 is the partial nucleotide sequence and deduced amino acid sequence for human GDF-8.

SEQ ID NO: 8 is the deduced partial amino acid sequence for human GDF-8.

SEQ ID NO: 9 is the amino acid sequence for primer SJL141.

SEQ ID NO: 10 is the amino acid sequence for primer SJL147.

15 SEQ ID NO: 11 is the nucleotide and deduced amino acid sequence for murine GDF-8.

SEQ ID NO: 12 is the deduced amino acid sequence for murine GDF-8.

SEQ ID NO: 13 is the nucleotide and deduced amino acid sequence for human GDF-8.

SEQ ID NO: 14 is the deduced amino acid sequence for human GDF-8.

5 SEQ ID NO's: 15 and 16 are nucleotide sequences for primer #83 and #84, respectively, which were used to map human GDF-8 in human/rodent somatic cell hybrids.

SEQ ID NO:17 is the nucleotide sequence of oligonucleotide #100 which corresponds to a sequence internal to the region flanked by primer #83 and #84.

-38-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: THE JOHNS HOPKINS UNIVERSITY
- 5 (ii) TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-8
- (iii) NUMBER OF SEQUENCES: 17
- (iv) CORRESPONDENCE ADDRESS:
- 10 (A) ADDRESSEE: Spensley Horn Jubas & Lubitz  
(B) STREET: 1880 Century Park East - Suite 500  
(C) CITY: Los Angeles  
(D) STATE: California  
(E) COUNTRY: USA  
(F) ZIP: 90067
- 15 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- 20 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER: PCT  
(B) FILING DATE: 18-MAR-1994  
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
- 25 (A) NAME: Wetherell, Jr., Ph.D., John R.,  
(B) REGISTRATION NUMBER: 31,678  
(C) REFERENCE/DOCKET NUMBER: FD-3413 CIP PCT
- (ix) TELECOMMUNICATION INFORMATION:
- 30 (A) TELEPHONE: (619) 455-5100  
(B) TELEFAX: (619) 455-5110

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs



-39-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 5 (vii) IMMEDIATE SOURCE:  
(B) CLONE: SJL141

- (ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 1..35  
10 (D) OTHER INFORMATION: /mod\_base= i  
/note= ""B" is defined as "I" (inosine)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCGGAATTCCG GBTGGVANRA YTGGRBRTB KCBCC  
35

- 15 (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vii) IMMEDIATE SOURCE:  
(B) CLONE: SJL147

- (ix) FEATURE:  
25 (A) NAME/KEY: CDS  
(B) LOCATION: 1..33

- (ix) FEATURE:  
(A) NAME/KEY: modified\_base  
(B) LOCATION: 1..33  
30 (D) OTHER INFORMATION: /mod\_base= i  
/note= ""B" is defined as "I" (inosine)"

-40-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGGGAATTCTC CABSCRCARC TNTCBACBRY CAT  
33

(2) INFORMATION FOR SEQ ID NO:3:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:  
(B) CLONE: ACM13

(ix) FEATURE:  
(A) NAME/KEY: CDS  
15 (B) LOCATION: 1..32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGCGGATCCA GAAGTCAAGG TGACAGACAC AC  
32

(2) INFORMATION FOR SEQ ID NO:4:

- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:  
(B) CLONE: ACM14

(ix) FEATURE:

-41-

(A) NAME/KEY: CDS  
(B) LOCATION: 1..33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGCGGATCCT CCTCATGAGC ACCCAGCG GTC  
5 33

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 550 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: mouse GDF-8

15 (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 59..436

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTAAGGTAGG AAGGATTTC A GGCTCTATTT ACATAATTGT TCTTTCCTTT TCACACAG  
20 58

AAT CCC TTT TTA GAA GTC AAG GTG ACA GAC ACA CCC AAG AGG TCC CGG  
106

Asn Pro Phe Leu Glu Val Lys Val Thr Asp Thr Pro Lys Arg Ser Arg  
1 5 10 15

25 AGA GAC TTT GGG CTT GAC TGC GAT GAG CAC TCC ACG GAA TCC CGG TGC  
154

Arg Asp Phe Gly Leu Asp Cys Asp Glu His Ser Thr Glu Ser Arg Cys  
20 25 30

30 TGC CGC TAC CCC CTC ACG GTC GAT TTT GAA GCC TTT GGA TGG GAC TGG  
202

-42-

Cys Arg Tyr Pro Leu Thr Val Asp Phe Glu Ala Phe Gly Trp Asp Trp  
                   35                                  40                                  45  
  
 ATT ATC GCA CCC AAA AGA TAT AAG GCC AAT TAC TGC TCA GGA GAG TGT  
 250  
 5 Ile Ile Ala Pro Lys Arg Tyr Lys Ala Asn Tyr Cys Ser Gly Glu Cys  
                   50                                  55                                  60  
  
 GAA TTT GTG TTT TTA CAA AAA TAT CCG CAT ACT CAT CTT GTG CAC CAA  
 298  
 10 Glu Phe Val Phe Leu Gln Lys Tyr Pro His Thr His Leu Val His Gln  
                   65                                  70                                  75                                  80  
  
 GCA AAC CCC AGA GGC TCA GCA GGC CCT TGC TGC ACT CCG ACA AAA ATG  
 346  
 Ala Asn Pro Arg Gly Ser Ala Gly Pro Cys Cys Thr Pro Thr Lys Met  
                                   85                                  90                                  95  
  
 15 TCT CCC ATT AAT ATG CTA TAT TTT AAT GGC AAA GAA CAA ATA ATA TAT  
 394  
 Ser Pro Ile Asn Met Leu Tyr Phe Asn Gly Lys Glu Gln Ile Ile Tyr  
                                   100                                  105                                  110  
  
 GGG AAA ATT CCA GCC ATG GTA GTA GAC CGC TGT GGG TGC TCA  
 436  
 20 Gly Lys Ile Pro Ala Met Val Val Asp Arg Cys Gly Cys Ser  
                   115                                  120                                  125  
  
 TGAGCTTTGC ATTAGGTTAG AAAGTTCCCA AGTCATGGAA GGTCTTCCCC TCAATTTCCA  
 496  
  
 25 AACTGTGAAT TCCTGCAGCC CGGGGGATCC ACTAGTTCTA GAGCGGCCGC CACC  
 550

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 126 amino acids  
       (B) TYPE: amino acid  
       (D) TOPOLOGY: linear  
  
 (ii) MOLECULE TYPE: protein  
  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

-43-

Asn Pro Phe Leu Glu Val Lys Val Thr Asp Thr Pro Lys Arg Ser Arg  
 1                      5                      10                      15  
 Arg Asp Phe Gly Leu Asp Cys Asp Glu His Ser Thr Glu Ser Arg Cys  
                     20                      25                      30  
 5 Cys Arg Tyr Pro Leu Thr Val Asp Phe Glu Ala Phe Gly Trp Asp Trp  
                     35                      40                      45  
 Ile Ile Ala Pro Lys Arg Tyr Lys Ala Asn Tyr Cys Ser Gly Glu Cys  
                     50                      55                      60  
 10 Glu Phe Val Phe Leu Gln Lys Tyr Pro His Thr His Leu Val His Gln  
                     65                      70                      75                      80  
 Ala Asn Pro Arg Gly Ser Ala Gly Pro Cys Cys Thr Pro Thr Lys Met  
                     85                      90                      95  
 Ser Pro Ile Asn Met Leu Tyr Phe Asn Gly Lys Glu Gln Ile Ile Tyr  
                     100                      105                      110  
 15 Gly Lys Ile Pro Ala Met Val Val Asp Arg Cys Gly Cys Ser  
                     115                      120                      125

## (2) INFORMATION FOR SEQ ID NO:7:

20 (i) SEQUENCE CHARACTERISTICS:  
     (A) LENGTH: 326 base pairs  
     (B) TYPE: nucleic acid  
     (C) STRANDEDNESS: single  
     (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (vii) IMMEDIATE SOURCE:  
     (B) CLONE: human GDF-8

(ix) FEATURE:  
     (A) NAME/KEY: CDS  
     (B) LOCATION: 3..326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

-44-

CA AAA AGA TCC AGA AGG GAT TTT GGT CTT GAC TGT GAT GAG CAC TCA  
 47  
 Lys Arg Ser Arg Arg Asp Phe Gly Leu Asp Cys Asp Glu His Ser  
 1 5 10 15

5 ACA GAA TCA CGA TGC TGT CGT TAC CCT CTA ACT GTG GAT TTT GAA GCT  
 95  
 Thr Glu Ser Arg Cys Cys Arg Tyr Pro Leu Thr Val Asp Phe Glu Ala  
 20 25 30

10 TTT GGA TGG GAT TGG ATT ATC GCT CCT AAA AGA TAT AAG GCC AAT TAC  
 143  
 Phe Gly Trp Asp Trp Ile Ile Ala Pro Lys Arg Tyr Lys Ala Asn Tyr  
 35 40 45

TGC TCT GGA GAG TGT GAA TTT GTA TTT TTA CAA AAA TAT CCT CAT ACT  
 191  
 15 Cys Ser Gly Glu Cys Glu Phe Val Phe Leu Gln Lys Tyr Pro His Thr  
 50 55 60

CAT CTG GTA CAC CAA GCA AAC CCC AGA GGT TCA GCA GGC CCT TGC TGT  
 239  
 His Leu Val His Gln Ala Asn Pro Arg Gly Ser Ala Gly Pro Cys Cys  
 20 65 70 75

ACT CCC ACA AAG ATG TCT CCA ATT AAT ATG CTA TAT TTT AAT GGC AAA  
 287  
 Thr Pro Thr Lys Met Ser Pro Ile Asn Met Leu Tyr Phe Asn Gly Lys  
 80 85 90 95

25 GAA CAA ATA ATA TAT GGG AAA ATT CCA GCG ATG GTA GTA  
 326  
 Glu Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met Val Val  
 100 105

## (2) INFORMATION FOR SEQ ID NO:8:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

-45-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Lys Arg Ser Arg Arg Asp Phe Gly Leu Asp Cys Asp Glu His Ser Thr  
 1 5 10 15  
 Glu Ser Arg Cys Cys Arg Tyr Pro Leu Thr Val Asp Phe Glu Ala Phe  
 5 20 25 30  
 Gly Trp Asp Trp Ile Ile Ala Pro Lys Arg Tyr Lys Ala Asn Tyr Cys  
 35 40 45  
 Ser Gly Glu Cys Glu Phe Val Phe Leu Gln Lys Tyr Pro His Thr His  
 50 55 60  
 10 Leu Val His Gln Ala Asn Pro Arg Gly Ser Ala Gly Pro Cys Cys Thr  
 65 70 75 80  
 Pro Thr Lys Met Ser Pro Ile Asn Met Leu Tyr Phe Asn Gly Lys Glu  
 85 90 95  
 15 Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met Val Val  
 100 105

## (2) INFORMATION FOR SEQ ID NO: 9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids  
 (B) TYPE: amino acid  
 20 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (vii) IMMEDIATE SOURCE:

(B) CLONE: SJL141

## 25 (ix) FEATURE:

- (A) NAME/KEY: Peptide  
 (B) LOCATION: 1..9  
 (D) OTHER INFORMATION: /note= "His = His, Asn, Lys, Asp or  
 Glu; Asp = Asp or Asn; Val = Val, Ile or Met; Ala  
 30 = Ala or Ser."

-46-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Trp His Asp Trp Val Val Ala Pro

1

5

(2) INFORMATION FOR SEQ ID NO:10:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:

(B) CLONE: SJL147

(ix) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..8

15

(D) OTHER INFORMATION: /note= "Ile = Ile, Val, Met, Thr or  
Ala; Asp = Asp or Glu; Gly = Gly or Ala."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Ile Val Asp Ser Cys Gly Cys

20

1

5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2676 base pairs

(B) TYPE: nucleic acid

25

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: Murine GDF-8



-47-

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 104..1231

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

5 GTCTCTCGGA CGGTACATGC ACTAATATTT CACTTGGCAT TACTCAAAAG CAAAAAGAAG  
 60  
 AAATAAGAAC AAGGGAACAAA AAAAGATTGT GCTGATTTTT AAA ATG ATG CAA AAA  
 115  
 Met Met Gln Lys  
 10 1  
 CTG CAA ATG TAT GTT TAT ATT TAC CTG TTC ATG CTG ATT GCT GCT GGC  
 163  
 Leu Gln Met Tyr Val Tyr Ile Tyr Leu Phe Met Leu Ile Ala Ala Gly  
 5 10 15 20  
 15 CCA GTG GAT CTA AAT GAG GGC AGT GAG AGA GAA GAA AAT GTG GAA AAA  
 211  
 Pro Val Asp Leu Asn Glu Gly Ser Glu Arg Glu Glu Asn Val Glu Lys  
 25 30 35  
 GAG GGG CTG TGT AAT GCA TGT GCG TGG AGA CAA AAC ACG AGG TAC TCC  
 20 259  
 Glu Gly Leu Cys Asn Ala Cys Ala Trp Arg Gln Asn Thr Arg Tyr Ser  
 40 45 50  
 AGA ATA GAA GCC ATA AAA ATT CAA ATC CTC AGT AAG CTG CGC CTG GAA  
 307  
 25 Arg Ile Glu Ala Ile Lys Ile Gln Ile Leu Ser Lys Leu Arg Leu Glu  
 55 60 65  
 ACA GCT CCT AAC ATC AGC AAA GAT GCT ATA AGA CAA CTT CTG CCA AGA  
 355  
 Thr Ala Pro Asn Ile Ser Lys Asp Ala Ile Arg Gln Leu Leu Pro Arg  
 30 70 75 80  
 GCG CCT CCA CTC CGG GAA CTG ATC GAT CAG TAC GAC GTC CAG AGG GAT  
 403  
 Ala Pro Pro Leu Arg Glu Leu Ile Asp Gln Tyr Asp Val Gln Arg Asp  
 85 90 95 100

-48-

GAC AGC AGT GAT GGC TCT TTG GAA GAT GAC GAT TAT CAC GCT ACC ACG  
 451  
 Asp Ser Ser Asp Gly Ser Leu Glu Asp Asp Asp Tyr His Ala Thr Thr  
 105 110 115

5 GAA ACA ATC ATT ACC ATG CCT ACA GAG TCT GAC TTT CTA ATG CAA GCG  
 499  
 Glu Thr Ile Ile Thr Met Pro Thr Glu Ser Asp Phe Leu Met Gln Ala  
 120 125 130

10 GAT GGC AAG CCC AAA TGT TGC TTT TTT AAA TTT AGC TCT AAA ATA CAG  
 547  
 Asp Gly Lys Pro Lys Cys Cys Phe Phe Lys Phe Ser Ser Lys Ile Gln  
 135 140 145

15 TAC AAC AAA GTA GTA AAA GCC CAA CTG TGG ATA TAT CTC AGA CCC GTC  
 595  
 Tyr Asn Lys Val Val Lys Ala Gln Leu Trp Ile Tyr Leu Arg Pro Val  
 150 155 160

20 AAG ACT CCT ACA ACA GTG TTT GTG CAA ATC CTG AGA CTC ATC AAA CCC  
 643  
 Lys Thr Pro Thr Thr Val Phe Val Gln Ile Leu Arg Leu Ile Lys Pro  
 165 170 175 180

ATG AAA GAC GGT ACA AGG TAT ACT GGA ATC CGA TCT CTG AAA CTT GAC  
 691  
 Met Lys Asp Gly Thr Arg Tyr Thr Gly Ile Arg Ser Leu Lys Leu Asp  
 185 190 195

25 ATG AGC CCA GGC ACT GGT ATT TGG CAG AGT ATT GAT GTG AAG ACA GTG  
 739  
 Met Ser Pro Gly Thr Gly Ile Trp Gln Ser Ile Asp Val Lys Thr Val  
 200 205 210

30 TTG CAA AAT TGG CTC AAA CAG CCT GAA TCC AAC TTA GGC ATT GAA ATC  
 787  
 Leu Gln Asn Trp Leu Lys Gln Pro Glu Ser Asn Leu Gly Ile Glu Ile  
 215 220 225

35 AAA GCT TTG GAT GAG AAT GGC CAT GAT CTT GCT GTA ACC TTC CCA GGA  
 835  
 Lys Ala Leu Asp Glu Asn Gly His Asp Leu Ala Val Thr Phe Pro Gly  
 230 235 240

-49-

CCA GGA GAA GAT GGG CTG AAT CCC TTT TTA GAA GTC AAG GTG ACA GAC  
883  
Pro Gly Glu Asp Gly Leu Asn Pro Phe Leu Glu Val Lys Val Thr Asp  
245 250 255 260

5 ACA CCC AAG AGG TCC CGG AGA GAC TTT GGG CTT GAC TGC GAT GAG CAC  
931  
Thr Pro Lys Arg Ser Arg Arg Asp Phe Gly Leu Asp Cys Asp Glu His  
265 270 275

10 TCC ACG GAA TCC CGG TGC TGC CGC TAC CCC CTC ACG GTC GAT TTT GAA  
979  
Ser Thr Glu Ser Arg Cys Cys Arg Tyr Pro Leu Thr Val Asp Phe Glu  
280 285 290

15 GCC TTT GGA TGG GAC TGG ATT ATC GCA CCC AAA AGA TAT AAG GCC AAT  
1027  
Ala Phe Gly Trp Asp Trp Ile Ile Ala Pro Lys Arg Tyr Lys Ala Asn  
295 300 305

20 TAC TGC TCA GGA GAG TGT GAA TTT GTG TTT TTA CAA AAA TAT CCG CAT  
1075  
Tyr Cys Ser Gly Glu Cys Glu Phe Val Phe Leu Gln Lys Tyr Pro His  
310 315 320

ACT CAT CTT GTG CAC CAA GCA AAC CCC AGA GGC TCA GCA GGC CCT TGC  
1123  
Thr His Leu Val His Gln Ala Asn Pro Arg Gly Ser Ala Gly Pro Cys  
325 330 335 340

25 TGC ACT CCG ACA AAA ATG TCT CCC ATT AAT ATG CTA TAT TTT AAT GGC  
1171  
Cys Thr Pro Thr Lys Met Ser Pro Ile Asn Met Leu Tyr Phe Asn Gly  
345 350 355

30 AAA GAA CAA ATA ATA TAT GGG AAA ATT CCA GCC ATG GTA GTA GAC CGC  
1219  
Lys Glu Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met Val Val Asp Arg  
360 365 370

35 TGT GGG TGC TCA TGAGCTTTGC ATTAGGTTAG AAACCTCCCA AGTCATGGAA  
1271  
Cys Gly Cys Ser  
375

GGTCTTCCCC TCAATTTCGA AACTGTGAAT TCAAGCACCA CAGGCTGTAG GCCTTGAGTA  
1331

TGCTCTAGTA ACGTAAGCAC AAGCTACAGT GTATGAACTA AAAGAGAGAA TAGATGCAAT  
1391

5 GGT TGGCATT CAACCACCAA AATAAACCAT ACTATAGGAT GTTGTATGAT TTCCAGAGTT  
1451

TTTGAAATAG ATGGAGATCA AATTACATTT ATGTCCATAT ATGTATATTA CAACTACAAT  
1511

10 CTAGGCAAGG AAGTGAGAGC ACATCTTGTG GTCTGCTGAG TTAGGAGGGT ATGATTAAAA  
1571

GGTAAAGTCT TATTTCTTAA CAGTTTCACT TAATATTTAC AGAAGAATCT ATATGTAGCC  
1631

TTTGTAAGT GTAGGATTGT TATCATTTAA AAACATCATG TACACTTATA TTTGTATTGT  
1691

15 ATACTTGGTA AGATAAAATT CCACAAAGTA GGAATGGGGC CTCACATACA CATTGCCATT  
1751

CCTATTATAA TTGGACAATC CACCACGGTG CTAATGCAGT GCTGAATGGC TCCTACTGGA  
1811

20 CCTCTCGATA GAACACTCTA CAAAGTACGA GTCTCTCTCT CCCTTCCAGG TGCATCTCCA  
1871

CACACACAGC ACTAAGTGTT CAATGCATTT TCTTTAAGGA AAGAAGAATC TTTTTTTCTA  
1931

GAGGTCAACT TTCAGTCAAC TCTAGCACAG CGGGAGTGAC TGCTGCATCT TAAAAGGCAG  
1991

25 CCAAACAGTA TTCATTTTTT AATCTAAATT TCAAAATCAC TGTCTGCCTT TATCACATGG  
2051

CAATTTTGTG GTAAAATAAT GCAAATGACT GGTTCATCA ATATTGTATA AAAGACTCTG  
2111

30 AAACAATTAC ATTTATATAA TATGTATACA ATATTGTTTT GTAAATAAGT GTCTCCTTTT  
2171

-51-

ATATTACTT TGGTATATTT TTACACTAAT GAAATTTCAG ATCATTAAAG TACAAAGACA  
 2231  
 TGTGATGTAT CACAAAAAAG GTGACTGCTT CTATTTTCAG GTGAATTAGC AGATTCAATA  
 2291  
 5 GTGGTCTTAA AACTCTGTAT GTTAAGATTA GAAGGTTATA TTACAATCAA TTTATGTATT  
 2351  
 TTTTACATTA TCAACTTATG GTTTCATGGT GGCTGTATCT ATGAATGTGG CTCCCAGTCA  
 2411  
 10 AATTTCATG CCCCAACCATT TTAAAAATTA CAAGCATTAC TAAACATACC AACATGTATC  
 2471  
 TAAAGAAATA CAAATATGGT ATCTCAATAA CAGCTACTTT TTTATTTTAT AATTGACAA  
 2531  
 TGAATACATT TCTTTTATTT ACTTCAGTTT TATAAATTGG AACTTTGTTT ATCAAATGTA  
 2591  
 15 TTGTAATCAT AGCTAAATGA AATTATTTCT TACATAAAAA TGTGTAGAAA CTATAAATTA  
 2651  
 AAGTGTTTTC ACATTTTGA AAGGC  
 2676

## (2) INFORMATION FOR SEQ ID NO:12:

- 20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 376 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: protein  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met	Met	Gln	Lys	Leu	Gln	Met	Tyr	Val	Tyr	Ile	Tyr	Leu	Phe	Met	Leu
1				5				10				15			
Ile	Ala	Ala	Gly	Pro	Val	Asp	Leu	Asn	Glu	Gly	Ser	Glu	Arg	Glu	Glu
			20				25					30			

-52-

	Asn	Val	Glu	Lys	Glu	Gly	Leu	Cys	Asn	Ala	Cys	Ala	Trp	Arg	Gln	Asn	
			35					40					45				
	Thr	Arg	Tyr	Ser	Arg	Ile	Glu	Ala	Ile	Lys	Ile	Gln	Ile	Leu	Ser	Lys	
			50				55					60					
5	Leu	Arg	Leu	Glu	Thr	Ala	Pro	Asn	Ile	Ser	Lys	Asp	Ala	Ile	Arg	Gln	
		65				70					75					80	
	Leu	Leu	Pro	Arg	Ala	Pro	Pro	Leu	Arg	Glu	Leu	Ile	Asp	Gln	Tyr	Asp	
					85					90					95		
10	Val	Gln	Arg	Asp	Asp	Ser	Ser	Asp	Gly	Ser	Leu	Glu	Asp	Asp	Asp	Tyr	
				100					105						110		
	His	Ala	Thr	Thr	Glu	Thr	Ile	Ile	Thr	Met	Pro	Thr	Glu	Ser	Asp	Phe	
				115				120					125				
	Leu	Met	Gln	Ala	Asp	Gly	Lys	Pro	Lys	Cys	Cys	Phe	Phe	Lys	Phe	Ser	
		130					135					140					
15	Ser	Lys	Ile	Gln	Tyr	Asn	Lys	Val	Val	Lys	Ala	Gln	Leu	Trp	Ile	Tyr	
		145				150					155					160	
	Leu	Arg	Pro	Val	Lys	Thr	Pro	Thr	Thr	Val	Phe	Val	Gln	Ile	Leu	Arg	
					165					170					175		
20	Leu	Ile	Lys	Pro	Met	Lys	Asp	Gly	Thr	Arg	Tyr	Thr	Gly	Ile	Arg	Ser	
			180						185					190			
	Leu	Lys	Leu	Asp	Met	Ser	Pro	Gly	Thr	Gly	Ile	Trp	Gln	Ser	Ile	Asp	
		195						200					205				
	Val	Lys	Thr	Val	Leu	Gln	Asn	Trp	Leu	Lys	Gln	Pro	Glu	Ser	Asn	Leu	
		210					215					220					
25	Gly	Ile	Glu	Ile	Lys	Ala	Leu	Asp	Glu	Asn	Gly	His	Asp	Leu	Ala	Val	
		225				230					235					240	
	Thr	Phe	Pro	Gly	Pro	Gly	Glu	Asp	Gly	Leu	Asn	Pro	Phe	Leu	Glu	Val	
					245					250					255		
30	Lys	Val	Thr	Asp	Thr	Pro	Lys	Arg	Ser	Arg	Arg	Asp	Phe	Gly	Leu	Asp	
				260					265					270			

-53-

Cys Asp Glu His Ser Thr Glu Ser Arg Cys Cys Arg Tyr Pro Leu Thr  
 275 280 285  
 Val Asp Phe Glu Ala Phe Gly Trp Asp Trp Ile Ile Ala Pro Lys Arg  
 290 295 300  
 5 Tyr Lys Ala Asn Tyr Cys Ser Gly Glu Cys Glu Phe Val Phe Leu Gln  
 305 310 315 320  
 Lys Tyr Pro His Thr His Leu Val His Gln Ala Asn Pro Arg Gly Ser  
 325 330 335  
 10 Ala Gly Pro Cys Cys Thr Pro Thr Lys Met Ser Pro Ile Asn Met Leu  
 340 345 350  
 Tyr Phe Asn Gly Lys Glu Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met  
 355 360 365  
 Val Val Asp Arg Cys Gly Cys Ser  
 370 375

15 (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2743 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Human GDF-8

25

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 59..1183

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAGAAAAGTA AAAGGAAGAA ACAAGAACAA GAAAAAAGAT TATATTGATT TAAAATC  
58

-54-

ATG CAA AAA CTG CAA CTC TGT GTT TAT ATT TAC CTG TTT ATG CTG ATT  
 106  
 Met Gln Lys Leu Gln Leu Cys Val Tyr Ile Tyr Leu Phe Met Leu Ile  
 1 5 10 15

5 GTT GCT GGT CCA GTG GAT CTA AAT GAG AAC AGT GAG CAA AAA GAA AAT  
 154  
 Val Ala Gly Pro Val Asp Leu Asn Glu Asn Ser Glu Gln Lys Glu Asn  
 20 25 30

10 GTG GAA AAA GAG GGG CTG TGT AAT GCA TGT ACT TGG AGA CAA AAC ACT  
 202  
 Val Glu Lys Glu Gly Leu Cys Asn Ala Cys Thr Trp Arg Gln Asn Thr  
 35 40 45

AAA TCT TCA AGA ATA GAA GCC ATT AAG ATA CAA ATC CTC AGT AAA CTT  
 250  
 15 Lys Ser Ser Arg Ile Glu Ala Ile Lys Ile Gln Ile Leu Ser Lys Leu  
 50 55 60

CGT CTG GAA ACA GCT CCT AAC ATC AGC AAA GAT GTT ATA AGA CAA CTT  
 298  
 20 Arg Leu Glu Thr Ala Pro Asn Ile Ser Lys Asp Val Ile Arg Gln Leu  
 65 70 75 80

TTA CCC AAA GCT CCT CCA CTC CGG GAA CTG ATT GAT CAG TAT GAT GTC  
 346  
 Leu Pro Lys Ala Pro Pro Leu Arg Glu Leu Ile Asp Gln Tyr Asp Val  
 85 90 95

25 CAG AGG GAT GAC AGC AGC GAT GGC TCT TTG GAA GAT GAC GAT TAT CAC  
 394  
 Gln Arg Asp Asp Ser Ser Asp Gly Ser Leu Glu Asp Asp Asp Tyr His  
 100 105 110

30 GCT ACA ACG GAA ACA ATC ATT ACC ATG CCT ACA GAG TCT GAT TTT CTA  
 442  
 Ala Thr Thr Glu Thr Ile Ile Thr Met Pro Thr Glu Ser Asp Phe Leu  
 115 120 125

ATG CAA GTG GAT GGA AAA CCC AAA TGT TGC TTC TTT AAA TTT AGC TCT  
 490  
 35 Met Gln Val Asp Gly Lys Pro Lys Cys Cys Phe Phe Lys Phe Ser Ser  
 130 135 140



-55-

AAA ATA CAA TAC AAT AAA GTA GTA AAG GCC CAA CTA TGG ATA TAT TTG  
 538  
 Lys Ile Gln Tyr Asn Lys Val Val Lys Ala Gln Leu Trp Ile Tyr Leu  
 145 150 155 160

5 AGA CCC GTC GAG ACT CCT ACA ACA GTG TTT GTG CAA ATC CTG AGA CTC  
 586  
 Arg Pro Val Glu Thr Pro Thr Thr Val Phe Val Gln Ile Leu Arg Leu  
 165 170 175

10 ATC AAA CCT ATG AAA GAC GGT ACA AGG TAT ACT GGA ATC CGA TCT CTG  
 634  
 Ile Lys Pro Met Lys Asp Gly Thr Arg Tyr Thr Gly Ile Arg Ser Leu  
 180 185 190

AAA CTT GAC ATG AAC CCA GGC ACT GGT ATT TGG CAG AGC ATT GAT GTG  
 682  
 15 Lys Leu Asp Met Asn Pro Gly Thr Gly Ile Trp Gln Ser Ile Asp Val  
 195 200 205

AAG ACA GTG TTG CAA AAT TGG CTC AAA CAA CCT GAA TCC AAC TTA GGC  
 730  
 20 Lys Thr Val Leu Gln Asn Trp Leu Lys Gln Pro Glu Ser Asn Leu Gly  
 210 215 220

ATT GAA ATA AAA GCT TTA GAT GAG AAT GGT CAT GAT CTT GCT GTA ACC  
 778  
 Ile Glu Ile Lys Ala Leu Asp Glu Asn Gly His Asp Leu Ala Val Thr  
 225 230 235 240

25 TTC CCA GGA CCA GGA GAA GAT GGG CTG AAT CCG TTT TTA GAG GTC AAG  
 826  
 Phe Pro Gly Pro Gly Glu Asp Gly Leu Asn Pro Phe Leu Glu Val Lys  
 245 250 255

30 GTA ACA GAC ACA CCA AAA AGA TCC AGA AGG GAT TTT GGT CTT GAC TGT  
 874  
 Val Thr Asp Thr Pro Lys Arg Ser Arg Arg Asp Phe Gly Leu Asp Cys  
 260 265 270

GAT GAG CAC TCA ACA GAA TCA CGA TGC TGT CGT TAC CCT CTA ACT GTG  
 922  
 35 Asp Glu His Ser Thr Glu Ser Arg Cys Cys Arg Tyr Pro Leu Thr Val  
 275 280 285

-56-

GAT TTT GAA GCT TTT GGA TGG GAT TGG ATT ATC GCT CCT AAA AGA TAT  
 970  
 Asp Phe Glu Ala Phe Gly Trp Asp Trp Ile Ile Ala Pro Lys Arg Tyr  
 290 295 300

5 AAG GCC AAT TAC TGC TCT GGA GAG TGT GAA TTT GTA TTT TTA CAA AAA  
 1018  
 Lys Ala Asn Tyr Cys Ser Gly Glu Cys Glu Phe Val Phe Leu Gln Lys  
 305 310 315 320

10 TAT CCT CAT ACT CAT CTG GTA CAC CAA GCA AAC CCC AGA GGT TCA GCA  
 1066  
 Tyr Pro His Thr His Leu Val His Gln Ala Asn Pro Arg Gly Ser Ala  
 325 330 335

15 GGC CCT TGC TGT ACT CCC ACA AAG ATG TCT CCA ATT AAT ATG CTA TAT  
 1114  
 Gly Pro Cys Cys Thr Pro Thr Lys Met Ser Pro Ile Asn Met Leu Tyr  
 340 345 350

20 TTT AAT GGC AAA GAA CAA ATA ATA TAT GGG AAA ATT CCA GCG ATG GTA  
 1162  
 Phe Asn Gly Lys Glu Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met Val  
 355 360 365

GTA GAC CGC TGT GGG TGC TCA TGAGATTTAT ATTAAGCGTT CATAACTTCC  
 1213  
 Val Asp Arg Cys Gly Cys Ser  
 370 375

25 TAAAACATGG AAGGTTTTCC CCTCAACAAT TTTGAAGCTG TGAAATTAAG TACCACAGGC  
 1273

TATAGGCCTA GAGTATGCTA CAGTCACTTA AGCATAAGCT ACAGTATGTA AACTAAAAGG  
 1333

30 GGGAATATAT GCAATGGTTG GCATTTAACC ATCCAAACAA ATCATACAAG AAAGTTTAT  
 1393

GATTTCCAGA GTTTTTCAGC TAGAAGGAGA TCAAATTACA TTTATGTTCC TATATATTAC  
 1453

AACATCGGCG AGGAAATGAA AGCGATTCTC CTTGAGTTCT GATGAATTAA AGGAGTATGC  
 1513

-57-

TTTAAAGTCT ATTTCTTTAA AGTTTTGTTT AATATTTACA GAAAAATCCA CATACAGTAT  
1573

TGGTAAATG CAGGATTGTT ATATACCATC ATTCGAATCA TCCTTAAACA CTTGAATTIA  
1633

5 TATTGTATGG TAGTATACTT GGTAAAGATAA AATTCACAA AAATAGGGAT GGTGCAGCAT  
1693

ATGCAATTTT CATTCTATT ATAATTGACA CAGTACATTA ACAATCCATG CCAACGGTGC  
1753

10 TAATACGATA GGCTGAATGT CTGAGGCTAC CAGGTTTATC ACATAAAAAA CATTCACTAA  
1813

AATAGTAAGT TTCTCTTTTTC TTCAGGTGCA TTTTCCTACA CCTCCAAATG AGGAATGGAT  
1873

TTTCTTTAAT GTAAGAAGAA TCATTTTTCT AGAGGTTGGC TTTCAATTCT GTAGCATACT  
1933

15 TGGAGAAACT GCATTATCTT AAAAGGCAGT CAAATGGTGT TTGTTTTTAT CAAAATGTCA  
1993

AAATAACATA CTGGGAGAAG TATGTAATTT TGTCTTTGGA AAATTACAAC ACTGCCTTTG  
2053

20 CAACACTGCA GTTTTTATGG TAAATAATA GAAATGATCG ACTCTATCAA TATTGTATAA  
2113

AAAGACTGAA ACAATGCATT TATATAATAT GTATACAATA TTGTTTTGTA AATAAGTGTC  
2173

TCCTTTTTTA TTTACTTTGG TATATTTTTA CACTAAGGAC ATTTCAAATT AAGTACTAAG  
2233

25 GCACAAAGAC ATGTCATGCA TCACAGAAAA GCAACTACTT ATATTTGAGA GCAAATTAGC  
2293

AGATTAAATA GTGGTCTTAA AACTCCATAT GTTAATGATT AGATGTTTAT ATTACAATCA  
2353

30 TTTTATATTT TTTTACATGA TTAACATTCA CTTATGGATT CATGATGGCT GTATAAAGTG  
2413

-58-

AATTTGAAAT TTCAATGGTT TACTGTCATT GTGTTTAAAT CTCAACGTTT CATTATTTTA  
2473

ATACTTGCAA AAACATTACT AAGTATACCA AAATAATTGA CTCTATTATC TGAAATGAAG  
2533

5 AATAAACTGA TGCTATCTCA ACAATAACTG TTACTTTTAT TTTATAATTT GATAATGAAT  
2593

ATATTTCTGC ATTTATTTAC TTCTGTTTTG TAAATTGGGA TTTTGTTAAT CAAATTTATT  
2653

10 GTACTATGAC TAAATGAAAT TATTTCTTAC ATCTAATTG TAGAAACAGT ATAAGTTATA  
2713

TTAAAGTGTT TTCACATTTT TTTGAAAGAC  
2743

(2) INFORMATION FOR SEQ ID NO:14:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 375 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

20 Met Gln Lys Leu Gln Leu Cys Val Tyr Ile Tyr Leu Phe Met Leu Ile  
1 5 10 15

Val Ala Gly Pro Val Asp Leu Asn Glu Asn Ser Glu Gln Lys Glu Asn  
20 25 30

25 Val Glu Lys Glu Gly Leu Cys Asn Ala Cys Thr Trp Arg Gln Asn Thr  
35 40 45

Lys Ser Ser Arg Ile Glu Ala Ile Lys Ile Gln Ile Leu Ser Lys Leu  
50 55 60

Arg Leu Glu Thr Ala Pro Asn Ile Ser Lys Asp Val Ile Arg Gln Leu  
65 70 75 80

-59-

	Leu	Pro	Lys	Ala	Pro	Pro	Leu	Arg	Glu	Leu	Ile	Asp	Gln	Tyr	Asp	Val	
					85					90					95		
	Gln	Arg	Asp	Asp	Ser	Ser	Asp	Gly	Ser	Leu	Glu	Asp	Asp	Asp	Tyr	His	
					100				105					110			
5	Ala	Thr	Thr	Glu	Thr	Ile	Ile	Thr	Met	Pro	Thr	Glu	Ser	Asp	Phe	Leu	
					115				120					125			
	Met	Gln	Val	Asp	Gly	Lys	Pro	Lys	Cys	Cys	Phe	Phe	Lys	Phe	Ser	Ser	
					130				135					140			
10	Lys	Ile	Gln	Tyr	Asn	Lys	Val	Val	Lys	Ala	Gln	Leu	Trp	Ile	Tyr	Leu	
					145				150					155		160	
	Arg	Pro	Val	Glu	Thr	Pro	Thr	Thr	Val	Phe	Val	Gln	Ile	Leu	Arg	Leu	
						165				170					175		
	Ile	Lys	Pro	Met	Lys	Asp	Gly	Thr	Arg	Tyr	Thr	Gly	Ile	Arg	Ser	Leu	
					180				185					190			
15	Lys	Leu	Asp	Met	Asn	Pro	Gly	Thr	Gly	Ile	Trp	Gln	Ser	Ile	Asp	Val	
					195				200					205			
	Lys	Thr	Val	Leu	Gln	Asn	Trp	Leu	Lys	Gln	Pro	Glu	Ser	Asn	Leu	Gly	
					210				215					220			
20	Ile	Glu	Ile	Lys	Ala	Leu	Asp	Glu	Asn	Gly	His	Asp	Leu	Ala	Val	Thr	
					225				230				235			240	
	Phe	Pro	Gly	Pro	Gly	Glu	Asp	Gly	Leu	Asn	Pro	Phe	Leu	Glu	Val	Lys	
					245					250					255		
	Val	Thr	Asp	Thr	Pro	Lys	Arg	Ser	Arg	Arg	Asp	Phe	Gly	Leu	Asp	Cys	
					260				265					270			
25	Asp	Glu	His	Ser	Thr	Glu	Ser	Arg	Cys	Cys	Arg	Tyr	Pro	Leu	Thr	Val	
					275				280					285			
	Asp	Phe	Glu	Ala	Phe	Gly	Trp	Asp	Trp	Ile	Ile	Ala	Pro	Lys	Arg	Tyr	
					290				295				300				
30	Lys	Ala	Asn	Tyr	Cys	Ser	Gly	Glu	Cys	Glu	Phe	Val	Phe	Leu	Gln	Lys	
					305				310				315			320	

-60-

Tyr Pro His Thr His Leu Val His Gln Ala Asn Pro Arg Gly Ser Ala  
325 330 335

Gly Pro Cys Cys Thr Pro Thr Lys Met Ser Pro Ile Asn Met Leu Tyr  
340 345 350

5 Phe Asn Gly Lys Glu Gln Ile Ile Tyr Gly Lys Ile Pro Ala Met Val  
355 360 365

Val Asp Arg Cys Gly Cys Ser  
370 375

(2) INFORMATION FOR SEQ ID NO:15:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:  
(B) CLONE: #83

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..34

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGCGGATCCG TGGATCTAAA TGAGAACAGT GAGC  
34

(2) INFORMATION FOR SEQ ID NO:16:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

-61-

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: #84

(ix) FEATURE:

5

(A) NAME/KEY: CDS

(B) LOCATION: 1..37

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGCGAATTCT CAGGTAATGA TTGTTTCCGT TGTAGCG  
37

10 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: #100

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

20 ACACTAAATC TTCAAGAATA  
20

**CLAIMS**

1. Substantially pure growth differentiation factor-8 (GDF-8) and functional fragments thereof.
2. An isolated polynucleotide sequence encoding the GDF-8 polypeptide of claim 1.
3. The polynucleotide of claim 2, wherein the GDF-8 nucleotide sequence is selected from the group consisting of the nucleic acid sequence of
  - a. FIGURE 5a, wherein T can also be U;
  - b. FIGURE 5b, wherein T can also be U;
  - 5 c. nucleic acid sequences complementary to FIGURE 5a;
  - d. nucleic acid sequences complementary to FIGURE 5b;
  - e. fragments of a. or c. that are at least 15 bases in length and that will selectively hybridize to genomic DNA which encodes the GDF-8 protein of FIGURE 5a; and
  - 10 f. fragments of b. or d. that are at least 15 bases in length and that will selectively hybridize to genomic DNA which encodes the GDF-8 protein of FIGURE 5b.
4. The polynucleotide sequence of claim 2, wherein the polynucleotide is isolated from a mammalian cell.
5. The polynucleotide of claim 4, wherein the mammalian cell is selected from the group consisting of mouse, rat, and human cell.
6. An expression vector including the polynucleotide of claim 2.
7. The vector of claim 6, wherein the vector is a plasmid.




8. The vector of claim 6, wherein the vector is a virus.
9. A host cell stably transformed with the vector of claim 6.
10. The host cell of claim 9, wherein the cell is prokaryotic.
11. The host cell of claim 9, wherein the cell is eukaryotic.
12. Antibodies reactive with the polypeptide of claim 1 or fragments thereof.
13. The antibodies of claim 12, wherein the antibodies are polyclonal.
14. The antibodies of claim 12, wherein the antibodies are monoclonal.
15. A method of detecting a cell proliferative disorder comprising contacting the antibody of claim 12 with a specimen of a subject suspected of having a GDF-8 associated disorder and detecting binding of the antibody.
16. The method of claim 15, wherein the cell is a muscle cell.
17. The method of claim 15, wherein the detecting is in vivo.
18. The method of claim 17, wherein the antibody is detectably labeled.
19. The method of claim 18, wherein the detectable label is selected from the group consisting of a radioisotope, a fluorescent compound, a bioluminescent compound and a chemiluminescent compound.
20. The method of claim 15, wherein the detection is in vitro.

21. The method of claim 20, wherein the antibody is detectably labeled.
22. The method of claim 21, wherein the label is selected from the group consisting of a radioisotope, a fluorescent compound, a bioluminescent compound, a chemoluminescent compound and an enzyme.
23. A method of treating a cell proliferative disorder associated with expression of GDF-8, comprising contacting the cells with a reagent which suppresses the GDF-8 activity.
24. The method of claim 23, wherein the reagent is an anti-GDF-8 antibody.
25. The method of claim 23, wherein the reagent is a GDF-8 antisense sequence.
26. The method of claim 23, wherein the cell is a muscle cell.
27. The method of claim 23, wherein the reagent which suppresses GDF-8 activity is introduced to a cell using a vector.
28. The method of claim 27, wherein the vector is a colloidal dispersion system.
29. The method of claim 28, wherein the colloidal dispersion system is a liposome.
30. The method of claim 29, wherein the liposome is essentially target specific.
31. The method of claim 30, wherein the liposome is anatomically targeted.

32. The method of claim 31, wherein the liposome is mechanistically targeted.
33. The method of claim 32, wherein the mechanistic targeting is passive.
34. The method of claim 32, wherein the mechanistic targeting is active.
35. The method of claim 34, wherein the liposome is actively targeted by coupling with a moiety selected from the group consisting of a sugar, a glycolipid, and a protein.
36. The method of claim 35, wherein the protein moiety is an antibody.
37. The method of claim 36, wherein the vector is a virus.
38. The method of claim 37, wherein the virus is an RNA virus.
39. The method of claim 38, wherein the RNA virus is a retrovirus.
40. The method of claim 39, wherein the retrovirus is essentially target specific.
41. The method of claim 40, wherein a moiety for target specificity is encoded by a polynucleotide inserted into the retroviral genome.
42. The method of claim 40, wherein a moiety for target specificity is selected from the group consisting of a sugar, a glycolipid, and a protein.
43. The method of claim 42, wherein the protein is an antibody.

1 / 15

HEART  
LUNG  
THYMUS  
BRAIN  
KIDNEY  
SEMINAL VESICLE  
PANCREAS  
INTESTINE  
SPLEEN  
TESTIS  
FAT  
UTERUS  
OVARY  
LIVER  
MUSCLE



— 2.9 kb

**FIG. 1**

SUBSTITUTE SHEET (RULE 26)

2 / 15

1 TTAAGGTAGGAAGGATTTTCAGGCTCTATTTACATAATTGTTCTTTTCTTTTCACACAGAA 60  
N  
61 TCCCTTTTITAGAAGTCAAGGTGACAGACACACCCAAGAGGTCCCGGAGAGACTTTGGGCT 120  
P F L E V K V T D T P [K R] S [R R] D F G L  
121 TGAAGCCTTTGGATGGGACTGGATTATCGACCCAAAAGATATAAGGCCAATTACTGCTC 180  
D C D E H S T E S R C C R Y P L T V D F  
181 TGAAGCCTTTGGATGGGACTGGATTATCGACCCAAAAGATATAAGGCCAATTACTGCTC 240  
E A F G W D W I I A P K R Y K A N Y C S  
241 AGGAGAGTGTGAATTTGTGTTTTTACAAAATATCCGCATACTCATCTTGTGCACCAAGC 300  
G E C E F V F L Q K Y P H T H L V H Q A  
301 AAACCCAGAGGCTCAGCAGGCCCTTGCTGCACTCCGACAAAATGTCTCCCATTAATAT 360  
N P R G S A G P C C T P T K M S P I N M  
361 GCTATATTTAATGGCAAAGAACAATAATATATGGGAAAATCCAGCCATGGTAGTAGA 420  
L Y F N G K E Q I I Y G K I P A M V V D  
421 CCGCTGTGGGTGCTCATGAGCTTTGCATTAGGTTAGAACTTCCAAGTCATGGAAGGTC 480  
R C G C S \*  
481 TTCCCTCAATTTCGAAACTGTGAATTCCTGCAGCCCGGGGATCCACTAGTTCTAGAGC 540  
541 GGCCGCCACC 550

FIG.2a

1 CAAAAGATCCAGAAGGGATTTTGGTCTTGAAGTGTGATGAGCACTCAACAGAATCAGAT 60  
[K R] S [R R] D F G L D C D E H S T E S R C  
61 GCTGTGCTTACCCTCTAACTGTGGATTTTGAAGCTTTTGGATGGGATTGGATTATCGCTC 120  
C R Y P L T V D F E A F G W D W I I A P  
121 CTAAAAGATATAAGGCCAATTACTGCTCTGGAGAGTGTGAATTTGTATTTTACAAAAT 180  
K R Y K A N Y C S G E C E F V F L Q K Y  
181 ATCCTCATACTCATCTGGTACACCAAGCAAACCCAGAGGTTACAGCAGGCCCTTGCTGTA 240  
P H T H L V H Q A N P R G S A G P C C T  
241 CTCCACAAAGATGTCTCCAATTAATATGCTATATTTAATGGCAAAGAACAATAATAT 300  
P T K M S P I N M L Y F N G K E Q I I Y  
301 ATGGGAAAATCCAGCGATGGTAGTA 326  
G K I P A M V V

FIG.2b

3/15

GDF-8 SRRDFGLDCDEHSTESRCCRYPLTVDF-EAFGWD-WI IAPKRYKANYCSGCECFVFLQKYP—  
 GDF-1 RPRRDAEPVLGGPGGACRARRLYVSF-REVGWHRWV IAPRGFLANYCQGCQALPVALSGSGCPP  
 BMP-2 REKROAKHKQRKRLKSSCKRHPLYVDF-SDVGWNDWI VAPPGYHAFYCHGECFPLADHLNS—  
 BMP-4 KRSPKHHSQRARKKNKNCRRHSLYVDF-SDVGWNDWI VAPPGYQAFYCHGDCPFPLADHLNS—  
 Vgr-1 SRGSGSSDYNGSELKTACKKHELIVSF-QDLGWQDWI IAPKGYAANYCDEGCSFPLNAHMA—  
 OP-1 LRMANVAENSSDQRQACKKHELIVSF-RDLGWQDWI IAPEGYAAYYCEGECFPLNSYMNA—  
 BMP-5 SRMSSVGDYNTSEQKQACKKHELIVSF-RDLGWQDWI IAPEGYAAYYCEGECFPLNAHMA—  
 BMP-3 EQTLKKARRKQWIEPRNCARRYLKVDF-ADIGWSEWI IAPSGYHANYCEGECPSHIACTSGSSL—  
 MIS GPGRAQRSAGATAADGPCALRELSVDL—RAERSVL IPE TYQANNCQGVCGWQSDRNPRI—  
 Inhibin $\alpha$  ALRLLQRPPEEPAAHANCHRVALNLSF-QELGWERWIVYPPSF IFHYCHGCGGLHIPNLSLPV—  
 Inhibin $\beta$ A HRRRRRGLECDGKV-NICCKKQFFVSF-KDIGWNDWI IAPSGYHANYCEGECPSHIACTSGSSL—  
 Inhibin $\beta$ B HRIRKRGLECDGRT-NLCCRQQFFIDF-RLIGWNDWI IAPTGYGNYCEGSCPAYLAGVPGSAS—  
 TGF- $\beta$ 1 HRRALDTNYCFSTTEKNCCVRQLYIDFRKDLGWK-WIHEPKGYHANFCLGPCPYIWSLD—  
 TGF- $\beta$ 2 KKRALDAAYCFRNVQDNCCLRPLYIDFRKDLGWK-WIHEPKGYANFCLGACPYLWSSD—  
 TGF- $\beta$ 3 KKRALDTNYCFRNLENCVVRPLYIDFRQDLGWK-WIHEPKGYANFCLGCPYLRSD—

GDF-8 -HTHLVHQANPRG—SAGPCCT-PTKMSPINMLYF-NGKEQIIYGKIPAMVVDRCCGS  
 GDF-1 ALNHAVLRALMHA—AAPGAADLPCCV—PARLSPISVLFF-DNSDNVVLROYEDMVVDECCOR  
 BMP-2 -TNHAI VQTLVNS—VNSKIPKACCV—PTELSAISMLYL-DENEKVVLKNYQDMVVEGCCOR  
 BMP-4 -TNHAI VQTLVNS—VNSSIPKACCV—PTELSAISMLYL-DEYDKVVLKNYQDMVVEGCCOR  
 Vgr-1 -TNHAI VQTLVHL—MNPEYVPKPCA—PTKLNALSVLYF-DDNSNVILKKYRNMVVRACGCH  
 OP-1 -TNHAI VQTLVHF—INPETVPKPCA—PTQLNALSVLYF-DDSSNVILKKYRNMVVRACGCH  
 BMP-5 -TNHAI VQTLVHL—MFPDHVPKPCA—PTKLNALSVLYF-DDSSNVILKKYRNMVVRACGCH  
 BMP-3 -NHATIQSIVRA-VGVVPGIPEPCCV—PEKMSSLSILFF-DENKNVVLKVYPNMTVE SCAOR  
 MIS -GNHVLLLLKMQA—RGAALARPPCCV—PTAYAGKLLISLSEER—ISAHVPMVMVATECCOR  
 Inhibin $\alpha$  -PGAPPTPAQPYS—LLPGAQPCCAALPGTMRPLHVRTTSDGGYSFKYETVFNLLTQHCACI  
 Inhibin $\beta$ A -SFHSTVINHYRMRGHSPFANLKSCCV—PTKLRPMSMLYY-DDGQNIKKDIGNMIVEECCGS  
 Inhibin $\beta$ B -SFHTAVVNQYRMRLNPGT-VNSCCI—PTKLSTMSMLYF-DDEYNI VKRDVPNMIVEECCGA  
 TGF- $\beta$ 1 -TQYSKVLALYNQ—HNPGASAAPCCV—PQALEPLPIVYV-VGRKPKV-EQLSNMIVRSCKCS  
 TGF- $\beta$ 2 -TQHSRVL SLYNT—LNPEASASPCCV—SQDLEPLTILYY-IGKTPKI-EQLSNMIVKSKCS  
 TGF- $\beta$ 3 -TTHSTVLGLYNT—LNPEASASPCCV—PQDLEPLTILYY-VGRTPKV-EQLSNMIVKSKCS

FIG.3

SUBSTITUTE SHEET (RULE 26)



1	GTCTCTCGGACGGTACATGCACTAATATTTTCTACTTGGCATTACTCAAAAGCAAAAAGAAG	60
61	AAATAAGAACAAAGGGAAAAAAAAGATTGTGCTGATTTTTAAATGATGCAAAAAGTCA	120
	M M Q K L Q	
121	AATGTATGTTTATATTTACCTGTTTCATGCTGATTGCTGCTGGCCAGTGGATCTAAATGA	180
	M Y V Y I Y L F M L I A A G P V D L N E	
181	GGGCAGTGAGAGAGAAGAAATGTGGAAGAGAGGGGCTGTGTAATGCATGTGCGTGGAG	240
	G S E R E E N V E K E G L C N A C A W R	
241	ACAAAACACGAGGTACTCCAGAATAGAAGCCATAAAATTCAAATCCTCAGTAAGCTGCC	300
	Q N T R Y S R I E A I K I Q I L S K L R	
301	CCTGGAACAGCTCCTAACATCAGCAAAGATGCTATAAGACAACCTTCTGCCAAGAGCGCC	360
	L E T A P N I S K D A I R Q L L P R A P	
361	TCCACTCCGGGAAGTGTGATCAGTACGACGTCCAGAGGGATGACAGCAGTGTGCTC	420
	P L R E L I D Q Y D V Q R D D S S D G S	
421	TTTGAAGATGACGATTATCAGCTACCACGGAAACAATCATTACCATGCCTACAGAGTC	480
	L E D D D Y H A T T E T I I T M P T E S	
481	TGACTTTCTAATGCAAGCGGATGGCAAGCCAAATGTTGCTTTTTAAATTTAGCTCTAA	540
	D F L M Q A D G K P K C C F F K F S S K	
541	AATACAGTACAACAAAGTAGTAAAGCCCAACTGTGGATATATCTCAGACCCGTCAAGAC	600
	I Q Y N K V V K A Q L W I Y L R P V K T	
601	TCCTACAACAGTGTTTGTGCAAAATCCTGAGACTCATCAACCCATGAAAGACGGTACAAG	660
	P T T V F V Q I L R L I K P M K D G T R	
661	GTATACTGGAATCCGATCTCTGAAACTTGACATGAGCCAGGCACTGGTATTGGCAGAG	720
	Y T G I R S L K L D M S P G T G I W Q S	
721	TATTGATGTGAAGACAGTGTGCAAAATGGCTCAAACAGCCTGAATCCAACCTTAGGCAT	780
	I D V K T V L Q N W L K Q P E S N L G I	
781	TGAAATCAAAGCTTTGGATGAGAATGGCCATGATCTTGCTGTAACTTCCCAGGACCAGG	840
	E I K A L D E N G H D L A V T F P G P G	
841	AGAAGATGGGCTGAATCCCTTTTTAGAAGTCAAGGTGACAGACACACCCAAGAGGTCCCG	900
	E D G L N P F L E V K V T D T P K R S R	
901	GAGAGACTTTGGGCTTGACTGCGATGAGCACTCCACGGAATCCCGGTGCTGCCGTACCC	960
	R D F G L D C D E H S T E S R C C R Y P	
961	CCTCACGGTCGATTTTGAAGCCTTTGGATGGGACTGGATTATCGCACCCAAAAGATATAA	1020
	L T V D F E A F G W D W I I A P K R Y K	
1021	GGCCAATTACTGCTCAGGAGAGTGTGAATTTGTGTTTTACAAAATATCCGCATACTCA	1080
	A N Y C S G E C E F V F L Q K Y P H T H	
1081	TCTTGTGCACCAAGCAAACCCAGAGGCTCAGCAGGCCCTTGCTGCACTCCGACAAAAAT	1140
	L V H Q A N P R G S A G P C C T P T K M	
1141	GTCTCCCATTAATATGCTATATTTTAAATGGCAAAGAACAATAATATATGGGAAAATTCC	1200
	S P I N M L Y F N G K E Q I I Y G K I P	
1201	AGCCATGGTAGTAGCCGTGTGGGTGCTCATGAGCTTTGCATTAGGTTAGAAACTTCCC	1260
	A M V V D R C G C S *	

**SUBSTITUTE SHEET (RULE 26)**



6 / 15

1261 AAGTCATGGAAGGTCTTCCCCTCAATTTGAAACTGTGAATTCAGCACACAGGCTGTA 1320  
1321 GGCCTTGAGTATGCTCTAGTAACGTAAGCACAAGCTACAGTGTGAACATAAGAGAGA 1380  
1381 ATAGATGCAATGGTTGGCATTCAACCACCAAAATAAACATACTATAGGATGTTGTATGA 1440  
1441 TTTCCAGAGTTTTTGAATAGATGGAGATCAAATTACATTTATGTCCATATATGTATATT 1500  
1501 ACAACTACAATCTAGGCAAGGAAGTGAGAGCACATCTTGTGGTCTGCTGAGTTAGGAGGG 1560  
1561 TATGATTAAGGTAAAGTCTTATTTCTAACAGTTTCACTTAATATTTACAGAAGAATC 1620  
1621 TATATGTAGCCTTTGTAAAGTGTAGGATTGTTATCATTTAAAAACATCATGTACACTTAT 1680  
1681 ATTTGTATTGTATACTTGGTAAGATAAAATTCACAAAGTAGGAATGGGGCCTCACATAC 1740  
1741 ACATTGCCATTCCATTATAATTGGACAATCCACCACGGTGCTAATGCAGTGTGAATGG 1800  
1801 CTCCTACTGGACCTCTCGATAGAACACTCTACAAAGTACGAGTCTCTCTCTCCCTCCAG 1860  
1861 GTGCATCTCCACACACACAGCACTAAGTGTCAATGCATTTTCTTTAAGGAAAGAAGT 1920  
1921 CTTTTTTTCTAGAGGTCAACTTTCAGTCAACTCTAGCACAGCGGAGTGAAGTGTGCATC 1980  
1981 TTTAAAGGCAGCCAAACAGTATTCATTTTTTAATCTAAATTTCAAAATCACTGTCTGCCT 2040  
2041 TTATCATGGCAATTTTGTGGTAAATAATGGAAATGACTGGTTCTATCAATATTGTAT 2100  
2101 AAAAGACTCTGAAACAATTACATTTATATAATATGTATACAATATTGTTTTGTAATAAG 2160  
2161 TGTCTCCTTTTATATTTACTTTGGTATATTTTACACTAATGAAATTTCAAATCATTAA 2220  
2221 GTACAAAGACATGTCATGTATCACAAAAAGGTGACTGCTTCTATTTTCAAGTGAATTAG 2280  
2281 CAGATTCAATAGTGGTCTTAAACTCTGTATGTTAAGATTAGAAGGTTATATTACAATCA 2340  
2341 ATTTATGTATTTTTTACATTATCAACTTATGGTTTCATGGTGGCTGTATCTATGAATGT 2400  
2401 GCTCCAGTCAAATTTCAATGCCCCACCATTTTAAAAATTACAAGCATTACTAACATAC 2460  
2461 CAACATGTATCTAAAGAAATACAAATATGGTATCTCAATAACAGCTACTTTTTTATTTA 2520  
2521 TAATTTGACAATGAATACATTTCTTTTATTTTACTTCAGTTTTATAAATTGGAACTTGTT 2580  
2581 TATCAAATGTATTGTACTCATAGCTAAATGAAATATTTCTTACATAAAAAATGTGTAGAA 2640  
2641 ACTATAAATTAAGTGTTCACATTTTGAAGGC 2676

FIG.5b

7/15

1 AAGAAAGCTAAAGGAAGAAACAAGAACAAGAAAAAGATTATATTGATTTTAAATCAT 60  
M  
61 GCAAAAGCTGCAACTCTGTGTTTATATTTACCTGTTTATGCTGATTGTTGCTGGTCCAGT 120  
Q K L Q L C V Y I Y L F M L I V A G P V  
121 GGATCTAAATGAGAACAGTGAGCAAAAAGAAATGTGGAAAAGAGGGGCTGTGTAATGC 180  
D L N E N S E Q K E N V E K E G L C N A  
181 ATGTACTTGGAGACAAAACACTAAATCTTCAAGAATAGAAGCCATTAAGATACAAATCCT 240  
C T W R Q N T K S S R I E A I K I Q I L  
241 CAGTAAACTTCGTCTGGAACAGCTCCTAACATCAGCAAAGATGTTATAAGACAACTTTT 300  
S K L R L E T A P N I S K D V I R Q L L  
301 ACCCAAAGCTCCTCCACTCCGGAACTGATTGATCAGTATGATGTCCAGAGGGATGACAG 360  
P K A P P L R E L I D Q Y D V Q R D D S  
361 CAGCGATGGCTCTTTGGAAGATGACGATTATCAGCTACAACGGAACAATCATTACCAT 420  
S D G S L E D D D Y H A T T E T I I T M  
421 GCCTACAGAGTCTGATTTTCTAATGCAAGTGGATGGAAAACCAAATGTTGCTTCTTTAA 480  
P T E S D F L M Q V D G K P K C C F F K  
481 ATTTAGCTCTAAAATACAATACAATAAAGTAGTAAAGGCCCAACTATGGATATATTTGAG 540  
F S S K I Q Y N K V V K A Q L W I Y L R  
541 ACCCGTCGAGACTCCTACAACAGTGTGTTGTCGAAATCTGAGACTCATCAAACCTATGAA 600  
P V E T P T T V F V Q I L R L I K P M K  
601 AGACGGTACAAGGTATACTGGAATCCGATCTCTGAAACTTGACATGAACCCAGGCACTGG 660  
D G T R Y T G I R S L K L D M N P G T G  
661 TATTTGGCAGAGCATTGATGTGAAGACAGTGTGCAAAATTGGCTCAAACAACCTGAATC 720  
I W Q S I D V K T V L Q N W L K Q P E S  
721 CAACTTAGGCATTGAAATAAAAGCTTTAGATGAGAATGGTCATGATCTTGCTGTAACTT 780  
N L G I E I K A L D E N G H D L A V T F  
781 CCCAGGACCAGGAGAAGATGGGCTGAATCCGTTTTAGAGGTCAAGGTAACAGACACACC 840  
P G P G E D G L N P F L E V K V T D T P  
841 AAAAAGATCCAGAAGGGATTTTGGTCTTGACTGTGATGAGCACTCAACAGAATCAGCATG 900  
K R S R R D F G L D C D E H S T E S R C  
901 CTGTCGTTACCTCTAACTGTGGATTTTGAAGCTTTTGGATGGGATTGGATTATCGCTCC 960  
C R Y P L T V D F E A F G W D W I I A P  
961 TAAAAGATATAAGGCCAATTACTGCTCTGGAGAGTGTGAATTTGTATTTTACAAAATA 1020  
K R Y K A N Y C S G E C E F V F L Q K Y  
1021 TCCTCATACTCATCTGGTACACCAAGCAAACCCAGAGGTTACAGAGGCCCTTGCTGTAC 1080  
P H T H L V H Q A N P R G S A G P C C T  
1081 TCCACAAAGATGTCTCCAATTAATATGCTATATTTAATGGCAAAGAACAATAATATA 1140  
P T K M S P I N M L Y F N G K E Q I I Y  
1141 TGGGAAAATTCCAGCGATGGTAGACCGCTGTGGGTGCTCATGAGATTTATATTAAGC 1200  
G K I P A M V V D R C G C S \*

FIG.5c

SUBSTITUTE SHEET (RULE 26)

1201 GTTCATAACTTCCTAAAACATGGAAGGTTTTCCCCTCAACAATTTGAAGCTGTGAAATT 1260  
1261 AAGTACCACAGGCTATAGGCCTAGAGTATGCTACAGTCACTTAAGCATAAGCTACAGTAT 1320  
1321 GTAAACTAAAAGGGGAATATATGCAATGGTTGGCATTAAACCATCCAAACAAATCATAC 1380  
1381 AAGAAAGTTTTATGATTTCCAGAGTTTTGAGCTAGAAGGAGATCAAATTACATTTATGT 1440  
1441 TCCTATATATTACAACATCGGCGAGGAAATGAAAGCGATTCTCCTTGAGTTCTGATGAAT 1500  
1501 TAAAGGAGTATGCTTTAAAGTCTATTTCTTTAAAGTTTTGTTTAATTTACAGAAAAAT 1560  
1561 CCACATACAGTATTGGTAAAATGCAGGATTGTTATATACCATCATTCGAATCATCCTTAA 1620  
1621 ACACTTGAATTTATATTGTATGGTAGTATACTGGTAAGATAAAATCCACAAAAATAGG 1680  
1681 GATGGTGCAGCATATGCAATTTCCATTCTATTATAATTGACACAGTACATTAACAATCC 1740  
1741 ATGCCAACGGTGCTAATACGATAGGCTGAATGCTGAGGCTACCAGGTTTATCACATAAA 1800  
1801 AAACATTCACTAAAATAGTAAGTTTCTTTTCTTCAGGTGCATTTTCTACACCTCCAA 1860  
1861 ATGAGGAATGCATTTTCTTTAATGTAAGAAGAATCATTTTTCTAGAGGTTGGCTTTCAAT 1920  
1921 TCTGTAGCATACTTGGAGAACTCCATTATCTTAAAGGCAGTCAAATGGTGTGTTGTTTT 1980  
1981 TATCAAAATGTCAAAATAACATACTTGGAGAAGTATGTAATTTTGTCTTTGGAAAATTAC 2040  
2041 AACACTGCCCTTTGCAACACTGCAGTTTTTATGGTAAAATAATAGAAATGATCGACTCTAT 2100  
2101 CAATATTGTATAAAAAGACTGAAACAATGCATTTATATAATATGTATACAATATTGTTTT 2160  
2161 GTAAATAAGTGTCTCCTTTTTTTATTTTACITTTGGTATATTTTTACACTAAGGACATTTCAA 2220  
2221 ATTAAGTACTAAGGCACAAAGACATGTCATGCATCACAGAAAAGCAACTACTTATATTTT 2280  
2281 AGAGCAAATTAGCAGATTAAATAGTGGTCTTAAACTCCATATGTTAATGATTAGATGGT 2340  
2341 TATATTACAATCATTTTATATTTTTTTACATGATTAACTTCACTTATGGATTCATGATG 2400  
2401 GCTGTATAAAGTGAATTTGAAATTTCAATGGTTTACTGTCATTGTGTTTAAATCTCAACG 2460  
2461 TTCCATTATTTAATACTTGCAAAAACATTACTAAGTATACCAAAATAATTGACTCTATT 2520  
2521 ATCTGAAATGAAGAATAAACTGATGCTATCTCAACAATACTGTTACTTTTATTTTATAA 2580  
2581 TTTGATAATGAATATATTTCTGCATTTATTTACTTCTGTTTTGTAAATTTGGGATTTTGT 2640  
2641 AATCAAATTTATTGTACTATGACTAAATGAAATTTTCTTACATCTAATTTGTAGAAAC 2700  
2701 AGTATAAGTTATATTAAAGTGTTCACATTTTTTTGAAAGAC 2743

FIG.5d

9/15

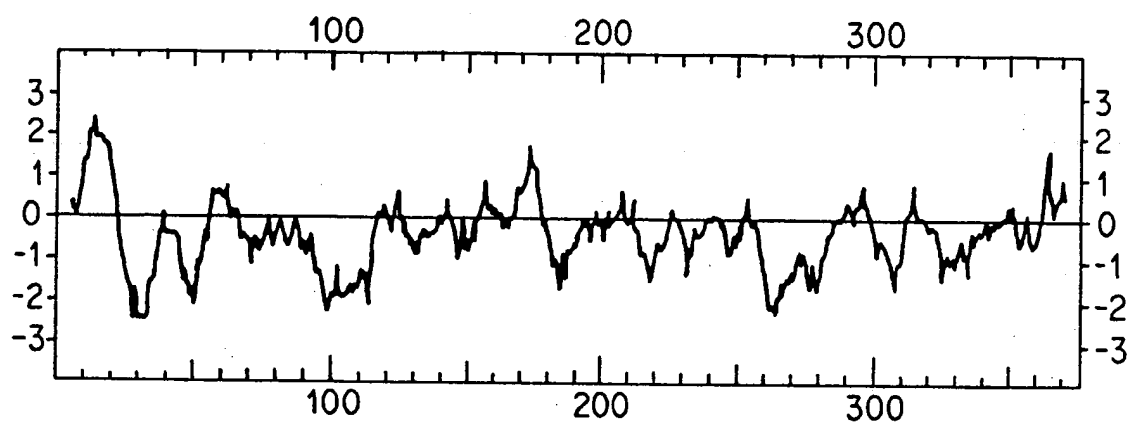


FIG. 6a

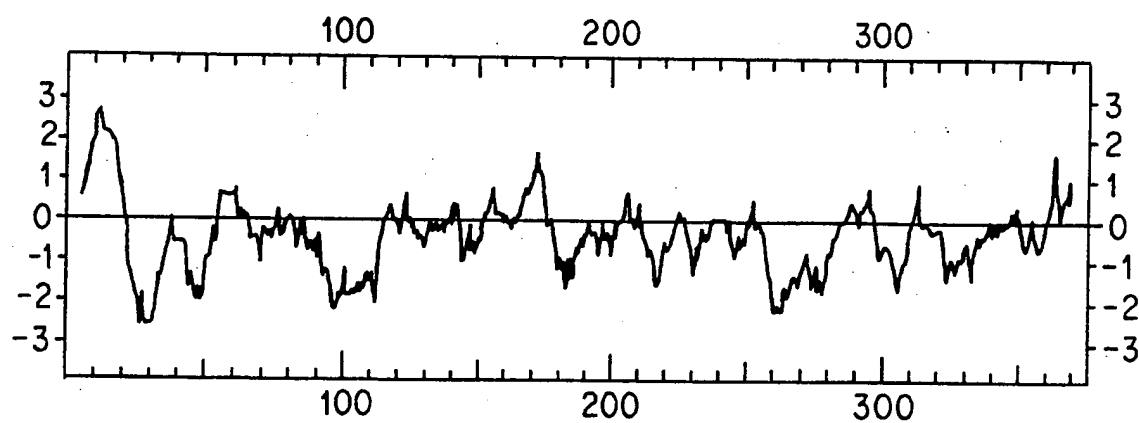


FIG. 6b

10/15

```

1  MMQKLQMYVYIYLFMLIAAGPVDLNEGSEREENVEKEGLCNACAWRONTR 50
   ||||| ||||| ||||| || ||||| ||||| |||||
1  MMQKLQLCVYIYLFMLIVAGPVDLNENSEQKENVEKEGLCNACTWRQNTK 49

51  YSRIEAIKIQILSKLRLETAPNISKDAIRQLLPRAPPLRELIDQYDVQRD 100
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
50  SSRIEAIKIQILSKLRLETAPNISKDVIRQLLPKAPPLRELIDQYDVQRD 99

101 DSSDGSLEDDDYHATTETIITMPTESDFLMQADGKPKCCFFKFSSKIQYN 150
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 DSSDGSLEDDDYHATTETIITMPTESDFLMQVDGKPKCCFFKFSSKIQYN 149

151 KVVKAQLWIYLRPVKTPTTVFVQILRLIKPMKDGTRYTGIRSLKLDMSGP 200
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 KVVKAQLWIYLRPVETPTTVFVQILRLIKPMKDGTRYTGIRSLKLDMPG 199

201 TGIWQSIDVKTVLQNLKQESNLGIEIKALDENGHD LAVTFPGPGEDGL 250
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 TGIWQSIDVKTVLQNLKQESNLGIEIKALDENGHD LAVTFPGPGEDGL 249

251 NPFLEVKVTDTPKRSRRDFGLDCDEHSTESRCCRYPLTVDFEAFGWDWI 300
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 NPFLEVKVTDTPKRSRRDFGLDCDEHSTESRCCRYPLTVDFEAFGWDWI 299

301 APKRYKANYCSGCECFVFLQKYPHTLVHQANPRGSAGPCCTPTKMSPIN 350
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 APKRYKANYCSGCECFVFLQKYPHTLVHQANPRGSAGPCCTPTKMSPIN 349

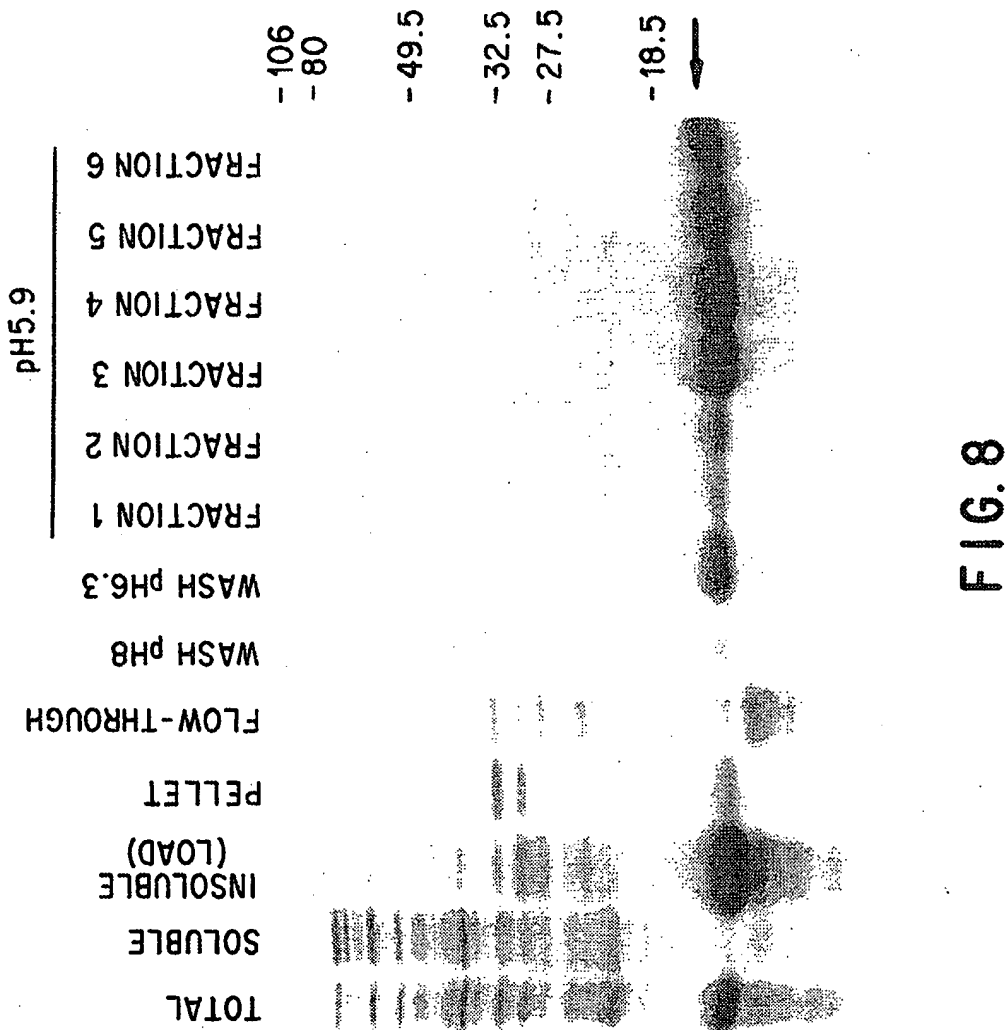
351 MLYFNGKEQIIYGKIPAMVDRCGCS 376
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 MLYFNGKEQIIYGKIPAMVDRCGCS 375

```

FIG.7

11 / 15

BEST AVAILABLE COPY



12/15

BEST AVAILABLE COPY

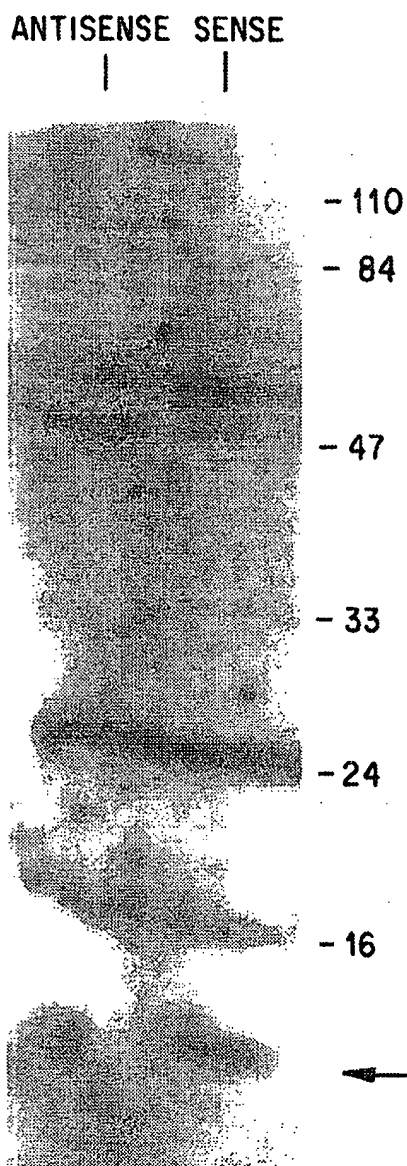


FIG. 9

13/15

BEST AVAILABLE COPY

FIG. 10

- 2.9 kb

UTERUS  
FAT  
OVARY  
LIVER  
MUSCLE  
TESTIS  
SPLEEN  
INTESTINE  
PANCREAS  
SEMINAL VESICLE  
KIDNEY  
BRAIN  
THYMUS  
LUNG  
HEART



14/15

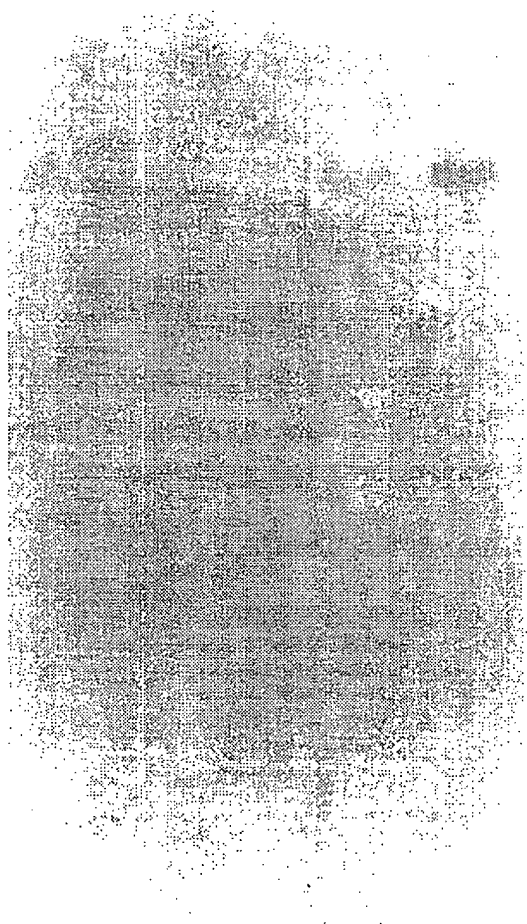
12.5 d PLACENTA

14.5 d PLACENTA

16.5 d PLACENTA

12.5 d EMBRYO

18.5 d EMBRYO



— 2.9 kb

BEST AVAILABLE COPY

FIG. 10b

15/15

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y M | H B1 CHO

1018 —  
506/517  
396  
344  
298

FIG. 11

BEST AVAILABLE COPY

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/03019

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C07K 13/00, 15/28; A61K 37/36; C12N 15/18, 15/10, 15/66

US CL : 530/399, 387.1; 536/23.5; 514/12; 435/69.1, 320.1, 252.3

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/399, 387.1; 536/23.5; 514/12; 435/69.1, 320.1, 252.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Molecular Endocrinology, VOLUME 6, NUMBER 11, issued 1992, "Isolation of Vgr-2, a Novel Member of the Transforming Growth Factor-beta-Related Gene Family", pages 1961-1968.	1-43
A	Proceedings of the National Academy of Sciences, VOLUME 88, issued May 1991, "Expression of growth/differentiation factor 1 in the nervous system: Conservation of a bicistronic structure", pages 4250-4254.	1-43
A	Molecular Endocrinology, VOLUME 4, NUMBER 7, issued 1990, "Identification of a Novel Member (GDF-1) of the Transforming Growth Factor-beta Superfamily", pages 1034-1040.	1-43

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be part of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

29 APRIL 1994

Date of mailing of the international search report

18 MAY 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

Shelly Guest Cermak

Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**International application No.  
PCT/US94/03019**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	The Journal of Biological Chemistry, VOLUME 268, NUMBER 5, issued 15 February 1993, "GDF-3 and GDF-9: Two New Members of the Transforming Growth Factor-beta Superfamily Containing a Novel Pattern of Cysteines", pages 3444-3449.	1-43

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/03019

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, Dialog: File Biochem. Medicine

search terms: Growth differentiation factor-8, GDF-8

Sequence Data: PIR, SwissPro, GenBank